Optimally adapted finite elements meshes

Jean-Marie Mirebeau

Optimal meshes for finite elements of arbitrary order

Jean-Marie Mirebeau

January 14, 2011

Abstract

Given a function f defined on a bounded domain $\Omega \subset \mathbb{R}^2$ and a number $N > 0$, we study the properties of the triangulation $\mathcal{T}_N$ that minimizes the distance between f and its interpolation on the associated finite element space, over all triangulations of at most N elements. The error is studied in the norm $X = L^p$ for $1 \leq p \leq \infty$ and we consider Lagrange finite elements of arbitrary polynomial degree $m-1$. We establish sharp asymptotic error estimates as $N \to +\infty$ when the optimal anisotropic triangulation is used, recovering the results on piecewise linear interpolation [3, 4, 12], an improving the results on higher degree interpolation [9, 10, 11]. These estimates involve invariant polynomials applied to the m-th order derivatives of f. In addition, our analysis also provides with practical strategies for designing meshes such that the interpolation error satisfies the optimal estimate up to a fixed multiplicative constant. We partially extend our results to higher dimensions for finite elements on simplicial partitions of a domain $\Omega \subset \mathbb{R}^d$.

Key words anisotropic finite elements, adaptive meshes, interpolation, nonlinear approximation.

AMS subject classifications 65D05, 65N15, 65N50

1 Introduction.

1.1 Optimal mesh adaptation

In finite element approximation, a usual distinction is between *uniform* and *adaptive* methods. In the latter, the elements defining the mesh may vary strongly in size and shape for a better adaptation to the local features of the approximated function f. This naturally raises the objective of characterizing and constructing an *optimal mesh* for a given function f.

Note that depending on the context, the function f may be fully known to us, either through an explicit formula or a discrete sampling, or observed through noisy measurements, or implicitly defined as the solution of a given partial differential equation.

In this paper, we assume that f is a function defined on a polygonal bounded domain $\Omega \subset \mathbb{R}^2$. For a given conforming triangulation $\mathcal{T}$ of Ω, and an arbitrary but fixed integer $m > 1$, we denote by $I_{m,\mathcal{T}}$ the standard interpolation operator on the Lagrange finite elements of degree $m-1$ space associated to $\mathcal{T}$. Given a norm X of interest and a number $N > 0$, the objective of finding the optimal mesh for f can be formulated as solving the optimization problem

$$\min_{\#(\mathcal{T}) \leq N} \|f - I_{m,\mathcal{T}} f\|_X,$$

where the minimum is taken over all conforming triangulations of cardinality N. We denote by $\mathcal{T}_N$ the minimizer of the above problem.

Our first objective is to establish sharp asymptotic error estimates that precisely describe the behavior of $\|f - I_{m,\mathcal{T}} f\|_X$ as $N \to +\infty$. Estimates of that type were obtained in [3, 4, 12] in the particular case of linear finite elements ($m - 1 = 1$) and with the error measured in $X = L^p$. They have the form

$$\limsup_{N \to +\infty} \Big(N \min_{\#(\mathcal{T}) \leq N} \|f - I_{m,\mathcal{T}} f\|_{L^p} \Big) \leq C \|\sqrt{|\det(d^2 f)|}\|_{L^\tau}, \quad \frac{1}{\tau} = \frac{1}{p} + 1, \tag{1}$$

which reveals that the convergence rate is governed by the quantity $\sqrt{|\det(d^2f)|}$, which depends nonlinearly the Hessian d^2f. This is heavily tied to the fact that we allow triangles with possibly highly anisotropic shape. In the present work, the polynomial degree $m-1$ is arbitrary and the quantities governing the convergence rate will therefore depend nonlinearly on the m-th order derivative $d^m f$.

Our second objective is to propose simple and practical ways of designing meshes which behave similar to the optimal one, in the sense that they satisfy the sharp error estimate up to a fixed multiplicative constant.

1.2 Main results and layout

We denote by $\mathbb{H}_m$ the space of homogeneous polynomials of degree m

$$\mathbb{H}_m := \mathrm{Span}\{x^k y^l \ : \ k+l=m\}.$$

For any triangle T, we denote by $I_{m,T}$ the local interpolation operator acting from $C^0(T)$ onto $\mathbb{P}_{m-1}$ the space of polynomials of total degree $m-1$. The image of $v \in C^0(T)$ by this operator is defined by the conditions

$$I_{m,T}v(\gamma) = v(\gamma),$$

for all points $\gamma \in T$ with barycentric coordinates in the set $\{0, \frac{1}{m-1}, \frac{2}{m-1}, \cdots, 1\}$. We denote by

$$e_{m,T}(v)_p := \|v - I_{m,T}v\|_{L^p(T)}$$

the interpolation error measured in the norm $L^p(T)$. We also denote by

$$e_{m,\mathcal{T}}(v)_p := \|v - I_{m,\mathcal{T}}v\|_{L^p} = \left(\sum_{T\in\mathcal{T}} e_{m,T}(v)_p^p\right)^{\frac{1}{p}},$$

the global interpolation error for a given triangulation $\mathcal{T}$, with the standard modification if $p=\infty$.

A key ingredient in this paper is a function defined by a *shape optimization problem*: for any fixed $1 \le p \le \infty$ and for any $\pi \in \mathbb{H}_m$, we define

$$K_{m,p}(\pi) := \inf_{|T|=1} e_{m,T}(\pi)_p. \tag{2}$$

Here, the infimum is taken over all triangles of area $|T|=1$. Note that from the homogeneity of π, we find that

$$\inf_{|T|=A} e_{m,T}(\pi)_p = K_{m,p}(\pi) A^{\frac{m}{2}+\frac{1}{p}}. \tag{3}$$

This optimization problem thus gives the shape of the triangles of a given area which is at best adapted to the polynomial π in the sense of minimizing the interpolation error measured in L^p. We refer to $K_{m,p}$ as the *shape function*. We discuss in §2 the main properties of this function.

Our asymptotic error estimate for the optimal triangulation is given by the following theorem.

Theorem 1.1 *For any polygonal domain $\Omega \subset \mathbb{R}^2$, and any function $f \in C^m(\Omega)$, there exists a sequence of triangulations $(\mathcal{T}_N)_{N\ge N_0}$, with $\#(\mathcal{T}_N) = N$ such that*

$$\limsup_{N\to\infty} N^{\frac{m}{2}} e_{m,\mathcal{T}_N}(f)_p \le \left\| K_{m,p}\left(\frac{d^m f}{m!}\right)\right\|_{L^q(\Omega)}, \quad \frac{1}{q} := \frac{m}{2}+\frac{1}{p} \tag{4}$$

An important feature of this estimate is the "lim sup" asymptotical operator. Recall that the upper limit of a sequence $(u_N)_{N\ge N_0}$ is defined by

$$\limsup_N u_N := \lim_{N\to\infty} \sup_{n\ge N} u_n,$$

and is in general strictly smaller than the supremum $\sup_{N\geq N_0} u_N$. It is still an open question to find an appropriate upper estimate for $\sup_N N^{m/2} e_{m,\mathcal{T}_N}(f)_p$ when optimally adapted anisotropic triangulations are used.

In the estimate (4), the m-th derivative $d^m f$ is identified to an homogeneous polynomial in $\mathbb{H}_m$:

$$\frac{d^m f}{m!} \sim \sum_{k+l=m} \frac{\partial^m f}{\partial^k x \partial^l y} \frac{x^k}{k!} \frac{y^l}{l!}.$$

In order to illustrate the sharpness of (4), we introduce a slight restriction on sequences of triangulations, following an idea in [3]: a sequence $(\mathcal{T}_N)_{N\geq N_0}$ of triangulations, such that $\#(\mathcal{T}_N) = N$, is said to be *admissible* if

$$\sup_{T\in\mathcal{T}_N} \mathrm{diam}(T) \leq C_A N^{-1/2}, \tag{5}$$

for some $C_A > 0$ independent of N. The following theorem shows that the estimate (4) cannot be improved when we restrict our attention to admissible sequences. It also shows that this class is reasonably large in the sense that (4) is ensured to hold up to small perturbation.

Theorem 1.2 *Let $\Omega \subset \mathbb{R}^2$ be a compact polygonal domain, and $f \in C^m(\Omega)$. Denote $\frac{1}{q} := \frac{m}{2} + \frac{1}{p}$. For all* admissible *sequences of triangulations $(\mathcal{T}_N)_{N\geq N_0}$, one has*

$$\liminf_{N\to\infty} N^{\frac{m}{2}} e_{m,\mathcal{T}_N}(f)_p \geq \left\| K_{m,p}\left(\frac{d^m f}{m!}\right)\right\|_{L^q(\Omega)}.$$

For all $\varepsilon > 0$, there exists an admissible *sequence of triangulations $(\mathcal{T}_N^\varepsilon)_{N\geq N_0}$, such that*

$$\limsup_{N\to\infty} N^{\frac{m}{2}} e_{m,\mathcal{T}_N^\varepsilon}(f)_p \leq \left\| K_{m,p}\left(\frac{d^m f}{m!}\right)\right\|_{L^q(\Omega)} + \varepsilon.$$

Note that the sequences $(\mathcal{T}_N^\varepsilon)_{N\geq N_0}$ satisfy the admissibility condition (5) with a constant $C_A(\varepsilon)$ which may explode as $\varepsilon \to 0$. The proofs of both theorems are given in §3. These proofs reveal that the construction of the optimal triangulation obeys two principles: (i) the triangulation should *equidistribute* the local approximation error $e_{m,T}(f)_p$ between each triangle and (ii) the aspect ratio of a triangle T should be *isotropic* with respect to a distorted metric induced by the local value of $d^m f$ on T (and therefore anisotropic in the sense of the euclidean metric). Roughly speaking, the quantity $\|K_{m,p}\left(\frac{d^m f}{m!}\right)\|_{L^q(T)}$ controls the local interpolation L^p-error estimate on a triangle T once this triangle is optimized with respect to the local properties of f. This type of estimate differs from those obtained in [2] which hold for any T, optimized or not, and involve the partial derivatives of f in a local coordinate system which is adapted to the shape of T.

The proof of the upper estimates in Theorem 1.2 involves the construction of an optimal mesh based on a patching strategy similar to [4]. However, inspection of the proof reveals that this construction becomes effective only when the number of triangles N becomes very large. Therefore it may not be useful in practical applications.

A more practical approach consists in deriving the above mentioned distorted metric from the exact or approximate data of $d^m f$, using the following procedure. To any $\pi \in \mathbb{H}_m$, we associate a symmetric positive definite matrix $h_\pi \in S_2^+$. If $z \in \Omega$ and $d^m f(z)$ is close to π, then the triangle T containing z should be isotropic in the metric h_π. The global metric is given at each point z by

$$h(z) = s(\pi_z) h_{\pi_z}, \quad \pi_z = d^m f(z),$$

where $s(\pi_z)$ is a scalar factor which depends on the desired accuracy of the finite element approximation. Once this metric has been properly identified, fast algorithms such as in [27, 26, 7] can be used to design a near-optimal mesh based on it. Recently in [20, 6], several algorithms have been rigorously proved to terminate and produce good quality meshes. Computing the map

$$\pi \in \mathbb{H}_m \mapsto h_\pi \in S_2^+, \tag{6}$$

is therefore of key use in applications. This problem is well understood in the case of linear elements ($m = 2$): the matrix h_π is then defined as the absolute value (in the sense of symmetric matrices) of the matrix associated to the quadratic form π. In contrast, the exact form of this map in the case $m \geq 3$ is not well understood.

In this paper, we propose algebraic strategies for computing the map (6) for $m = 3$ which corresponds to quadratic elements. These strategies have been implemented in an open-source Mathematica code [25]. In a similar manner, we address the algebraic computation of the shape function $K_{m,p}(\pi)$ from the coefficients of $\pi \in \mathbb{H}_m$, when $m \geq 3$. All these questions are addressed in §4, 5 and 6.

In §4, we discuss the particular case of linear ($m = 2$) and quadratic ($m = 3$) elements. In this case, it is possible to obtain explicit formulas for $K_{m,p}(\pi)$ from the coefficients of π. In the case $m = 2$, this formula is of the form

$$K_{2,p}(ax^2 + 2bxy + cy^2) = \sigma\sqrt{|b^2 - ac|},$$

where the constant σ only depends on p and the sign of $b^2 - ac$, and we therefore recover the known estimate (1) from Theorem 1.1. The formula for $m = 3$ involves the discriminant of the third degree polynomial $d^3 f$. Our analysis also leads to an algebraic computation of the map (6). We want to mention that a different strategy for the the construction of the distorted metric and the derivation of error estimate for finite element of arbitrary order was proposed in [9]. In this approach, the distorted metric is obtained at a point $z \in \Omega$ by finding the largest ellipse contained in a level set of the polynomial $d^m f_z$. This optimization problem has connections with the one that defines the shape function in (2) as we shall explain in §2. The approach proposed in the present work in the case $m = 3$ has the advantage of avoiding the use of numerical optimization, the metric being directly derived from the coefficients of $d^m f$.

In §5, we address the case $m > 3$. In this case, explicit formulas for $K_{m,p}(\pi)$ seem out of reach. However we can introduce explicit functions $\mathbf{K}_m(\pi)$ which are polynomials in the coefficients of π, and are equivalent to $K_{m,p}(\pi)$, leading therefore to similar asymptotic error estimates up to multiplicative constants. At the current stage, we did not obtain a simple solution to the algebraic computation of the map (6) in the case $m > 3$. The derivation of $\mathbf{K}_m$ is based on the theory of invariant polynomials due to Hilbert. Let us mention that this theory was also recently applied in [22] to image processing tasks such as affine invariant edge detection and denoising.

We finally discuss in §6 the possible extension of our analysis to simplicial elements in higher dimension. This extension is not straightforward except in the case of linear elements $m = 2$.

2 The shape function

In this section, we establish several properties of the function $K_{m,p}$ which will be of key use in the sequel. We assume that $m \geq 2$ is an integer, and $p \in [1, \infty]$. We equip the finite dimensional vector space $\mathbb{H}_m$ with a norm $\|\cdot\|$ defined as the supremum of the coefficients

$$\text{If } \pi(x, y) = \sum_{i=0}^{m} a_i x^i y^{m-i}, \text{ then } \|\pi\| = \max_{0 \leq i \leq m} |a_i|. \tag{7}$$

Our first result shows that the function $K_{m,p}$ vanishes on a set of polynomials which has a simple algebraic characterization.

Proposition 2.1 *We denote by $s_m := \lfloor \frac{m}{2} \rfloor + 1$ the smallest integer strictly larger than $m/2$. The vanishing set of $K_{m,p}$ is the set of polynomials which have a generalized root of multiplicity at least s_m:*

$$K_{m,p}(\pi) = 0 \Leftrightarrow \pi(x, y) = (\alpha x + \beta y)^{s_m} \tilde{\pi}, \text{ for some } \alpha, \beta \in \mathbb{R} \text{ and } \tilde{\pi} \in \mathbb{H}_{m-s_m}.$$

Proof: We denote by T_{eq} a fixed equilateral triangle of unit area, centered at 0.

We first assume that $\pi(x, y) = (\alpha x + \beta y)^{s_m} \tilde{\pi}$. Then there exists a rotation $R \in \mathcal{O}_2$ and $\hat{\pi} \in H_{m-s_m}$ such that

$$\pi \circ R(x, y) = x^{s_m} \hat{\pi}(x, y) = x^{s_m} \left(\sum_{i=0}^{m-s_m} a_i x^i y^{m-s_m-i} \right),$$

Therefore denoting by ϕ_ε the linear transform $\phi_\varepsilon(x,y) = R\left(\varepsilon x, \frac{y}{\varepsilon}\right)$ we obtain

$$\|\pi\circ\phi_\varepsilon\| = \max_{i=0,\cdots,m-s_m} |a_i|\varepsilon^{2s_m-m+2i} \le \varepsilon^{2s_m-m}\|\hat\pi\| \to 0 \text{ as } \varepsilon\to 0.$$

Consequently

$$e_{m,\phi_\varepsilon(T_{\rm eq})}(\pi)_p = e_{m,T_{\rm eq}}(\pi\circ\phi_\varepsilon)_p \to 0 \text{ as } \varepsilon\to 0.$$

Since $|\det\phi_\varepsilon| = 1$, the triangles $\phi_\varepsilon(T_{\rm eq})$ have unit area, and therefore $K_{m,p}(\pi) = 0$.

Conversely, let $\pi\in\mathbb{H}_m\backslash\{0\}$ be such that $K_{m,p}(\pi) = 0$. Then there exists a sequence $(T_n)_{n\ge 0}$ of triangles with unit area such that $e_{m,T_n}(\pi)_p \to 0$. We remark that the interpolation error $e_T(\pi)_p$ of $\pi\in\mathbb{H}_m$ is invariant by a translation $\tau_h : z\mapsto z+h$ of the triangle T. Indeed $\pi - \pi\circ\tau_h \in \mathbb{P}_{m-1}$ so that

$$\|\pi - I_{m,T}\pi\|_{L^p(\tau_h(T))} = \|\pi\circ\tau_h - I_{m,T}(\pi\circ\tau_h)\|_{L^p(T)} = \|\pi - I_{m,T}\pi\|_{L^p(T)}. \quad (8)$$

Hence we may assume that the barycenter of T_n is 0, and write $T_n = \phi_n(T_{\rm eq})$, for some linear transform ϕ_n with $\det\phi_n = 1$. Since $e_{m,T_{\rm eq}}(\cdot)_p$ is a norm on $\mathbb{H}_m$, it follows that $\pi\circ\phi_n\to 0$.

The linear transform ϕ_n has a singular value decomposition

$$\phi_n = U_n\circ D_n\circ V_n, \text{ where } U_n, V_n\in\mathcal{O}_2, \text{ and } D_n = \begin{pmatrix}\varepsilon_n & 0\\ 0 & 1/\varepsilon_n\end{pmatrix},\ 0<\varepsilon_n\le 1.$$

Since the orthogonal group $\mathcal{O}_2$ is compact, there is a uniform constant C such that

$$\|\pi\circ V\| \le C\|\pi\|,\ \ \pi\in\mathbb{H}_m,\ V\in\mathcal{O}_2.$$

Therefore

$$\|\pi\circ U_n\circ D_n\| = \|\pi\circ U_n\circ D_n\circ V_n\circ V_n^{-1}\| \le C\|\pi\circ\phi_n\| \to 0.$$

Denoting by $a_{i,n}$ the coefficient of x^iy^{m-i} in $\pi\circ U_n$, we find that $a_{i,n}\varepsilon_n^{2i-m}$ tends to 0 as $n\to+\infty$. In the case where $i<s_m$, this implies that $a_{i,n}$ tends to 0 as $n\to+\infty$

Moreover, again by compactness of $\mathcal{O}_2$, we may assume, up to a subsequence, that U_n converges to some $U\in\mathcal{O}_2$. Denoting by a_i the coefficient of x^iy^{m-i} in $\pi\circ U$, we thus find that $a_i = 0$ if $i<s_m$. This This implies that $\pi\circ U(x,y) = x^{s_m}\hat\pi(x,y)$ which concludes the proof. ⋄

Remark 2.1 *In the simple case $m=2$, we infer from Proposition 2.1 that $K_{2,p}(\pi) = 0$ if and only if π is of the form $\pi(x,y) = x^2$ up to a rotation, and therefore a one-dimensional function. For such a function, the optimal triangle T degenerates to a segment in the y direction, i.e. optimal triangles of a fixed area tend to be infinitely long in one direction. This situation also holds when $m>2$. Indeed, we see in the second part in the proof of Proposition 2.1 that if π is a non-trivial polynomial such that $K_{m,p}(\pi) = 0$, then ε_n must tends to 0 as $n\to+\infty$. This shows that $T_n = \phi_n(T)$ tends to be infinitely flat in the direction Ue_y with $e_y = (0,1)$. However, $K_{m,p}(\pi) = 0$ does not any longer mean that π is a polynomial of one variable.*

Our next result shows that the function $K_{m,p}$ is homogeneous, and obeys an invariance property with respect to linear change of variables.

Proposition 2.2 *For all $\pi\in\mathbb{H}_m$, $\lambda\in\mathbb{R}$ and $\phi\in\mathcal{L}(\mathbb{R}^2)$,*

$$K_{m,p}(\lambda\pi) = |\lambda|K_{m,p}(\pi) \quad (9)$$

$$K_{m,p}(\pi\circ\phi) = |\det\phi|^{m/2}K_{m,p}(\pi) \quad (10)$$

Proof: The homogeneity property (9) is a direct consequence of the definitions of $K_{m,p}$. In order to prove the invariance property (10) we assume in a first part that $\det\phi\neq 0$ and we define $\tilde T := \frac{\phi(T)}{\sqrt{|\det\phi|}}$ and $\tilde\pi(z) := \pi(\sqrt{|\det\phi|}z) = |\det\phi|^{m/2}\pi(z)$.

We now remark that the local interpolant $I_{m,T}$ commutes with linear change of variables in the sense that, when ϕ is an invertible linear transform,

$$I_{m,T}(v \circ \phi) = (I_{m,\phi(T)}v) \circ \phi, \tag{11}$$

for all continuous function v and triangle T. Using this commutation formula we obtain

$$\begin{aligned} e_{m,T}(\pi \circ \phi)_p &= |\det \phi|^{-1/p} e_{m,\phi(T)}(\pi)_p \\ &= e_{m,\tilde{T}}(\tilde{\pi})_p \\ &= |\det \phi|^{m/2} e_{m,\tilde{T}}(\pi)_p. \end{aligned}$$

Since the map $T \mapsto \tilde{T}$ is a bijection of the set of triangles onto itself, leaving the area invariant, we obtain the relation (10) when ϕ is invertible. When $\det \phi = 0$, the polynomial $\pi \circ \phi$ can be written $(\alpha x + \beta y)^m$ so that $K_{m,p}(P \circ \phi) = 0$ by Proposition 2.1. $\diamond$

The functions $K_{m,p}$ are not necessarily continuous, but the following properties will be sufficient for our purposes.

Proposition 2.3 *The function $K_{m,p}$ is upper semi-continuous in general, and continuous if $m = 2$ or m is odd. Moreover the following property holds:*

$$\text{If } \pi_n \to \pi \text{ and } K_{m,p}(\pi_n) \to 0 \text{ then } K_{m,p}(\pi) = 0. \tag{12}$$

Proof: The upper semi-continuity property comes from the fact that the infimum of a family of upper semi-continuous functions is an upper semi-continuous function. We apply this fact to the functions $\pi \mapsto e_{m,T}(\pi)_p$ indexed by triangles which are obviously continuous.

For any polynomial $\pi \in \mathbb{H}_2$, $\pi = ax^2 + 2bxy + cy^2$, we define $\det \pi = ac - b^2$. It will be shown in §4 that $K_{2,p}(\pi) = \sigma_p \sqrt{|\det \pi|}$, where σ_p only depends on the sign of $\det \pi$. This clearly implies the continuity of $K_{2,p}$. We next turn to the proof of the continuity of $K_{m,p}$ for odd m. Consider a polynomial $\pi \in \mathbb{H}_m$. If $K_{m,p}(\pi) = 0$ then the upper semi-continuity of $K_{m,p}$, combined with its non-negativity, implies that it is continuous at π. Otherwise, assume that $K_{m,p}(\pi) > 0$. Consider a sequence $\pi_n \in \mathbb{H}_m$ converging to π, and a sequence ϕ_n of linear transformations satisfying $\det \phi_n = 1$, and such that

$$\lim_{n \to +\infty} e_{\phi_n(T_{\rm eq})}(\pi_n) = \liminf_{\pi^* \to \pi} K_{m,p}(\pi^*) := \lim_{r \to 0} \inf_{\|\pi - \pi\| \le r} K_{m,p}(\pi^*).$$

If the sequence ϕ_n admits a converging subsequence $\phi_{n_k} \to \phi$, it follows that

$$K_{m,p}(\pi) \le e_{\phi(T_{\rm eq})}(\pi) = \lim_{k \to +\infty} e_{\phi_{n_k}(T_{\rm eq})}(\pi_{n_k}) = \liminf_{\pi^* \to \pi} K_{m,p}(\pi^*).$$

This asserts that $K_{m,p}$ is lower semi continuous at π, and therefore continuous at π since we already know that $K_{m,p}$ is upper semi-continuous.

If ϕ_n does not admit any converging subsequence, then we invoke the SVD decomposition $\phi_n = U_n \circ D_n \circ V_n$, where $U_n, V_n \in \mathcal{O}_2$ and $D_n = \text{diag}(\varepsilon_n, \frac{1}{\varepsilon_n})$, where $0 < \varepsilon_n \le 1$. (Here and below, we use the shorthand $\text{diag}(a, b)$ to denote the diagonal matrix with entries a and b) The compactness of $\mathcal{O}_2$ implies that U_n admits a converging subsequence $U_{n_k} \to U$. In particular $\pi_{n_k} \circ U_{n_k}$ converges to $\pi \circ U$. Therefore, denoting by $a_{i,n}$ the coefficient of $x^i y^{m-i}$ in $\pi_n \circ U_n$, the subsequence a_{i,n_k} converges to the coefficient a_i of $x^i y^{m-i}$ in $\pi \circ U$. Observe also that $\varepsilon_n \to 0$, otherwise some converging subsequence could be extracted from ϕ_n. Since $e_{\phi_n(T_{\rm eq})}(\pi_n) = e_{T_{\rm eq}}(\pi_n \circ \phi_n)$, the sequence of polynomials $\pi_n \circ \phi_n$ is uniformly bounded, and so is the sequence $\pi_n \circ U_n \circ D_n$. Therefore the sequences $(a_{i,n} \varepsilon_n^{2i-m})_{n \ge 0}$ are uniformly bounded. It follows that $a_i = 0$ when $i < \frac{m}{2}$. Since m is odd, this implies that $\pi \circ U(x, y) = x^{s_m} \tilde{\pi}(x, y)$ and Proposition 2.1 implies that $K_{m,p}(\pi) = 0$ which contradicts the hypothesis $K_{m,p}(\pi) > 0$.

Last, we prove property (12). The assumption $K_{m,p}(\pi_n) \to 0$ is equivalent to the existence of a sequence $T_n = \phi_n(T_{\rm eq})$ with $\det \phi_n = 1$ such that $e_{m,T_n}(\pi_n)_p \to 0$. Reasoning in a similar way as in the proof of Proposition 2.1, we first obtain that $\pi_n \circ \phi_n \to 0$, and we then invoke the SVD decomposition of ϕ_n to build a converging sequence of orthogonal matrices $U_n \to U$ and a sequence $0 < \varepsilon_n \le 1$ such that if $a_{i,n}$ is the coefficient of $x^i y^{m-i}$ in $\pi_n \circ U_n$, we have $a_{i,n}\varepsilon_n^{2i-m} \to 0$. When $i < s_m$, it follows that $a_{i,n} \to 0$ and therefore $\pi \circ U(x,y) = x^{s_m}\hat{\pi}(x,y)$. The result follows from Proposition 2.1. $\diamond$

We finally make a connection between the shape function and the approach developed in [9]. For all $\pi \in \mathbb{H}_m$, we denote by Λ_π as the level set of $|\pi|$ for the value 1,

$$\Lambda_\pi = \{(x,y) \in \mathbb{R}^2,\ |\pi(x,y)| \le 1\}. \tag{13}$$

We now define

$$K_m^{\mathcal{E}}(\pi) = \left(\sup_{E\in\mathcal{E},\ E\subset\Lambda_\pi} |E|\right)^{-m/2}. \tag{14}$$

where the supremum is taken over the set $\mathcal{E}$ of all ellipses centered at 0. The optimization problem defining $K_m^{\mathcal{E}}$ is equivalent to

$$\inf\{\det H \ :\ H \in S_2^+ \text{ and } \forall z \in \mathbb{R}^2, \langle Hz, z\rangle \ge |\pi(z)|^{2/m}\}, \tag{15}$$

where S_2^+ is the cone of 2×2 symmetric definite positive matrices. The minimizing ellipse E^* is then given by $\{\langle Hz, z\rangle \le 1\}$. The optimization problem described in (15) is quadratic in dimension 2, and subject to (infinitely many) linear constraints. This apparent simplicity is counterbalanced by the fact that it is non convex. In particular, it does not have unique solutions and may also have no solution.

Proposition 2.4 *On $\mathbb{H}_m$, one has the equivalence*

$$cK_m^{\mathcal{E}} \le K_{m,p} \le CK_m^{\mathcal{E}}$$

with constant $0 < c \le C$ independent of p.

Proof: Let $T_{\rm eq}$ denote an equilateral triangle of unit area, and B its circumscribed disk. It is easy to see that the inscribed disc is $B/2$.

We first show that $K_{m,p} \le CK_m^{\mathcal{E}}$. Let $\pi \in \mathbb{H}_m$, and let E_n be a sequence of ellipsoids inscribed in Λ_π and such that $|E_n|$ tends to $\sup\{|E| \ :\ E \in \mathcal{E},\ E \subset \Lambda_\pi\}$ as $n \to +\infty$. We write $E_n = \lambda_n\phi_n(B)$, where ϕ_n is a linear transform such that $\det\phi_n = 1$ and $\lambda_n > 0$. We define the triangle $T_n = \phi_n(T_{\rm eq})$ which satisfies $|T_n| = 1$. We then have

$$\begin{array}{rcl} K_{m,p}(\pi) & \le & \|\pi - I_{m,T_n}\pi\|_{L^p(T_n)} \\ & = & \|\pi\circ\phi_n - (I_{m,T_n}\pi)\circ\phi_n\|_{L^p(T_{\rm eq})} \\ & = & \|\pi\circ\phi_n - I_{m,T_{\rm eq}}(\pi\circ\phi_n)\|_{L^p(T_{\rm eq})} \\ & \le & \|\pi\circ\phi_n - I_{m,T_{\rm eq}}(\pi\circ\phi_n)\|_{L^\infty(T_{\rm eq})}, \end{array}$$

where we have used the commutation formula (11).

Remarking that $I_{m,T_{\rm eq}}$ is a continuous operator from $\mathbb{H}_m$ to $\mathbb{P}_{m-1}$ in the sense of any norm since these spaces are finite dimensional, we thus obtain

$$\begin{array}{rcl} K_{m,p}(\pi) & \le & C_1\|\pi\circ\phi_n\|_{L^\infty(T_{\rm eq})} \\ & \le & C_1\|\pi\circ\phi_n\|_{L^\infty(B)} \\ & = & C_1\|\pi\|_{L^\infty(\phi_n(B))} \\ & = & C_1\lambda_n^{-m}\|\pi\|_{L^\infty(E_n)} \\ & \le & C_1\left(\dfrac{|E_n|}{|B|}\right)^{-m/2}, \end{array}$$

where we have used the fact that $|\pi| \le 1$ in $E_n \subset \Lambda_\pi$. Letting $n \to +\infty$, we obtain that $K_{m,p}(\pi) \le CK_m^{\mathcal{E}}(\pi)$ with $C = C_1|B|^{m/2}$.

We next prove that $cK_m^{\mathcal{E}} \le K_{m,p}$. Let T_n be a sequence of triangles of unit area such that $e_{m,T_n}(\pi)_p$ tends to $K_{m,p}(\pi)$ as $n \to +\infty$. As already remarked in (8) the interpolation error is invariant by translation. We may therefore assume that the triangles T_n have their barycenter at the origin. Then there exists linear transforms ϕ_n with $\det \phi_n = 1$, such that $T_n = \phi_n(T_{\rm eq})$. We now write

$$\begin{aligned} \|\pi\|_{L^\infty(\phi_n(B/2))} &\leq \|\pi\|_{L^\infty(T_n)} \\ &= \|\pi \circ \phi_n\|_{L^\infty(T_{\rm eq})} \\ &\leq C_2 e_{m,T_{\rm eq}}(\pi \circ \phi)_1, \\ &\leq C_2 e_{m,T_{\rm eq}}(\pi \circ \phi)_p, \\ &= C_2 e_{m,T_n}(\pi)_p, \end{aligned}$$

where we have used the fact that $\|\cdot\|_{L^\infty(T_{\rm eq})}$ and $e_{T_{\rm eq}}(\cdot)_1$ are equivalent norms on $\mathbb{H}_m$, and that $e_{m,T_{\rm eq}}(\cdot)_p$ is an increasing function of p since $|T_{\rm eq}| = 1$. By homogeneity, it follows that if $\lambda_n := (C_2 e_{m,T_n}(\pi)_p)^{-1/m}$, we have

$$\|\pi\|_{L^\infty(\lambda_n \phi_n(B/2))} \leq \lambda_n^m C_2 e_{m,T_n}(\pi)_p = 1.$$

Therefore the ellipse $E_n := \lambda_n \phi_n(B/2)$ is contained in Λ_π, so that

$$K_m^{\mathcal{E}} \leq |E_n|^{-m/2} = \left(\lambda_n^2 \frac{|B|}{4}\right)^{-m/2} = C_2 \left(\frac{4}{|B|}\right)^{m/2} e_{m,T_n}(\pi)_p.$$

Letting $n \to +\infty$, we obtain that $cK_m^{\mathcal{E}} \leq K_{m,p}$ with $c = C_2^{-1} \left(\frac{|B|}{4}\right)^{m/2}$. ◇

Remark 2.2 *Since $K_{m,p}$ and $K_m^{\mathcal{E}}$ are equivalent, they must vanish on the same set, and therefore Proposition 2.1 is also valid for $K_m^{\mathcal{E}}$. It also easy to see that $K_m^{\mathcal{E}}$ satisfies the homogeneity and invariance properties stated for $K_{m,p}$ in (9) and (10), as well as the continuity properties stated in Proposition 2.3.*

Remark 2.3 *The continuity of the functions $K_{m,p}$ and $K_m^{\mathcal{E}}$ can be established when m is odd or equal to 2, as shown by Proposition 2.3, but seems to fail otherwise. In particular, direct computation shows that $K_4^{\mathcal{E}}(x^2y^2 - \varepsilon y^4)$ is independent of $\varepsilon > 0$ and strictly smaller than $K_4^{\mathcal{E}}(x^2y^2)$. Therefore $K_4^{\mathcal{E}}$ is upper semi-continuous but discontinuous at the point $x^2y^2 \in \mathbb{H}_4$.*

3 Optimal estimates

This section is devoted to the proofs of our main theorems, starting with the lower estimate of Theorem 1.2, and continuing with the upper estimates involved in both Theorem 1.1 and 1.2.

Throughout this section, for the sake of notational simplicity, we fix the parameters m and p and use the shorthand

$$K = K_{m,p} \text{ and } e_T(\pi) = e_{m,T}(\pi)_p.$$

For each point $z \in \Omega$ we define

$$\pi_z := \frac{d^m f_z}{m!} \in \mathbb{H}_m,$$

where $f \in C^m(\Omega)$ is the function in the statement of the theorems. We denote by

$$\omega(r) := \sup_{\|z-z'\| \leq r} \|\pi_z - \pi_{z'}\|,$$

the modulus of continuity of $z \mapsto \pi_z$ with the norm $\|\cdot\|$ defined by (7). Note that $\omega(r) \to 0$ as $r \to 0$.

3.1 Lower estimate

In this proof we will use an estimate by below of the local interpolation error.

Proposition 3.1 *Assume that $1 \leq p < \infty$. There exists a constant $C > 0$, depending on f and Ω, such that for all triangle $T \subset \Omega$ and $z \in T$,*

$$e_T(f)^p \geq K^p(\pi_z)|T|^{\frac{mp}{2}+1} - C(\operatorname{diam} T)^{mp}|T|\omega(\operatorname{diam} T). \tag{16}$$

Proof: Denoting by $\mu_z \in \mathbb{P}_m$ the Taylor development of f at the point z up to degree m, we obtain

$$f(z+u) - \mu_z(z+u) = m \int_{t=0}^{1} (\pi_{z+tu}(u) - \pi_z(u))(1-t)^{m-1} dt.$$

and therefore

$$\|f - \mu_z\|_{L^\infty(T)} \le C_0 \operatorname{diam}(T)^m \omega(\operatorname{diam}(T)),$$

where C_0 is a fixed constant. By construction π_z is the homogenous part of μ_z of degree m, and therefore $\mu_z - \pi_z \in \mathbb{P}_{m-1}$. It follows that for any triangle T, we have

$$\mu_z - I_{m,T}\mu_z = \pi_z - I_{m,T}\pi_z. \tag{17}$$

We therefore obtain

$$\begin{aligned}
|e_T(f) - e_T(\pi_z)| &\le \|(f - I_{m,T}f) - (\pi_z - I_{m,T}\pi_z)\|_{L^p(T)} \\
&\le |T|^{1/p}\|(f - I_{m,T}f) - (\mu_z - I_{m,T}\mu_z)\|_{L^\infty(T)} \\
&= |T|^{1/p}\|(I - I_{m,T})(f - \mu_z)\|_{L^\infty(T)} \\
&\le C_1|T|^{1/p}\|f - \mu_z\|_{L^\infty(T)} \\
&\le C_0C_1|T|^{1/p} \operatorname{diam}(T)^m \omega(\operatorname{diam}(T))
\end{aligned}$$

where C_1 is the norm of the operator $I - I_{m,T}$ in $L^\infty(T)$ which is independent of T.

From (3) we know that $e_T(\pi_z) \ge |T|^{\frac{m}{2}+\frac{1}{p}} K(\pi_z)$, and therefore

$$e_T(f) \ge K(\pi_z)|T|^{\frac{m}{2}+\frac{1}{p}} - C_0C_1|T|^{1/p} \operatorname{diam}(T)^m \omega(\operatorname{diam}(T)).$$

We now remark that for all $p \in [1, \infty)$ the function $r \mapsto r^p$ is convex, and therefore if a, b, c are positive numbers, and $a \ge b - c$ then $a^p \ge \max\{0, b-c\}^p \ge b^p - pcb^{p-1}$. Applying this to our last inequality we obtain

$$e_T(f)^p \ge K^p(\pi_z)|T|^{\frac{mp}{2}+1} - pC_0C_1(K(\pi_z))^{p-1}|T|^{(p-1)(\frac{m}{2}+\frac{1}{p})+\frac{1}{p}} \operatorname{diam}(T)^m \omega(\operatorname{diam} T).$$

Since $|T|^{(p-1)(\frac{m}{2}+\frac{1}{p})+\frac{1}{p}} = |T|^{(p-1)\frac{m}{2}}|T| \le (\operatorname{diam} T)^{m(p-1)}|T|$, this leads to

$$e_T(f)^p \ge K^p(\pi_z)|T|^{\frac{mp}{2}+1} - C(\operatorname{diam} T)^{mp}|T|\omega(\operatorname{diam} T),$$

where $C := pC_0C_1(\sup_{z\in\Omega} K(\pi_z))^{p-1}$. ◇

We now turn to the proof of the lower estimate in Theorem 1.2 in the case where $p < \infty$. Consider a sequence $(\mathcal{T}_N)_{N\ge N_0}$ of triangulations which is admissible in the sense of equation (5). Therefore, there exists a constant C_A such that

$$\operatorname{diam} T \le C_A N^{-1/2},\ N \ge N_0,\ T \in \mathcal{T}_N$$

For $T \in \mathcal{T}_N$, we combine this estimate with (16), which gives

$$e_T(f)^p \ge K^p(\pi_z)|T|^{\frac{mp}{2}+1} - (C_A N^{-1/2})^{mp}|T|C\omega(C_A N^{-1/2}).$$

Averaging over T, we obtain

$$e_T(f)^p \ge \int_T K^p(\pi_z)|T|^{\frac{mp}{2}} dz - |T|N^{-\frac{mp}{2}} C_A^{mp} C\omega(C_A N^{-1/2}).$$

Summing on all $T \in \mathcal{T}_N$, and denoting by T_z^N the triangle in $\mathcal{T}_N$ containing the point $z \in \Omega$, we obtain the estimate

$$e_{\mathcal{T}_N}(f)^p \ge \int_\Omega K(\pi_z)|T_z^N|^{\frac{mp}{2}} dz - N^{-\frac{mp}{2}}\varepsilon(N), \tag{18}$$

where $\varepsilon(N) := |\Omega| C_A^{mp} C\omega(C_A N^{-1/2}) \to 0$ as $N \to +\infty$. The function $z \mapsto |T_z^N|$ is linked with the number of triangles in the following way:

$$\int_\Omega \frac{dz}{|T_z^N|} = \sum_{T\in\mathcal{T}_N} \int_T \frac{1}{|T|} = N.$$

On the other hand, with $\frac{1}{q} = \frac{m}{2} + \frac{1}{p}$, we have by Hölder's inequality,

$$\int_\Omega K^q(\pi_z)dz \le \left(\int_\Omega K^p(\pi_z)|T_z^N|^{\frac{mp}{2}} dz\right)^{q/p} \left(\int_\Omega \frac{1}{|T_z^N|} dz\right)^{1-q/p}. \tag{19}$$

Combining the above, we obtain a lower bound for the integral term in (18) which is independent of $\mathcal{T}_N$:

$$\int_\Omega K^p(\pi_z)|T_z^N|^{\frac{mp}{2}} dz \ge \left(\int_\Omega K^q(\pi_z)dz\right)^{p/q} N^{-mp/2}.$$

Injecting this lower bound in (18) we obtain $e_{\mathcal{T}_N}(f)^p \ge \left[\left(\int_\Omega K^q(\pi_z)dz\right)^{p/q} - \varepsilon(N)\right] N^{-mp/2}$. This allows us to conclude

$$\liminf_{N\to+\infty} N^{\frac{m}{2}} e_{\mathcal{T}_N}(f) \ge \left(\int_\Omega K^q(\pi_z)dz\right)^{\frac{1}{q}}, \tag{20}$$

which is the desired estimate.

The case $p = \infty$ follows the same ideas. Adapting Proposition 3.1, one proves that

$$e_T(f) \ge K(\pi_z)|T|^{\frac{m}{2}} - C(\operatorname{diam} T)^m \omega(\operatorname{diam} T).$$

and therefore

$$e_{\mathcal{T}_N}(f) \ge \left\|K(\pi_z)|T_z^N|^{\frac{m}{2}}\right\|_{L^\infty(\Omega)} - N^{-\frac{m}{2}}\varepsilon(N), \tag{21}$$

where $\varepsilon(N) := C_A^m C\omega(C_A N^{-\frac{1}{2}}) \to 0$ as $N \to +\infty$. The Holder inequality now reads

$$\int_\Omega K(\pi_z)^{\frac{2}{m}} dz \le \left\|K(\pi_z)^{\frac{2}{m}} |T_z^N|\right\|_{L^\infty(\Omega)} \left\|\frac{1}{|T_z^N|}\right\|_{L^1(\Omega)}$$

equivalently

$$\left\|K(\pi_z)|T_z^N|^{\frac{m}{2}}\right\|_{L^\infty(\Omega)} \ge \left(\int_\Omega K(\pi_z)^{\frac{2}{m}} dz\right)^{\frac{m}{2}} N^{-\frac{m}{2}}.$$

Combining this with (21), this leads to the desired estimate (20) with $p = \infty$ and $q = \frac{2}{m}$.

Remark 3.1 *This proof reveals the two principles which characterize the optimal triangulations. Indeed, the lower estimate (20) becomes an equality only when both inequalities in (16) and (19) are equality. The first condition - equality in (16) - is met when each triangle T has an optimal shape, in the sense that $e_T(\pi_z) = K(\pi_z)|T|^{\frac{m}{2}+\frac{1}{p}}$ for some $z \in T$. The second condition - equality in (19) - is met when the ratio between $K^p(\pi_z)|T_z^N|^{\frac{mp}{2}}$ and $|T_z^N|^{-1}$ is constant, or equivalently $K(\pi_z)|T|^{\frac{m}{2}+\frac{1}{p}}$ is independent of the triangle T. Combined with the first condition, this means that the error $e_T(f)^p$ is equidistributed over the triangles, up to the perturbation by $(\operatorname{diam} T)^{mp}|T|\omega(\operatorname{diam} T)$ which becomes neglectible as N grows.*

3.2 Upper estimate

We first remark that the upper estimate in Theorem 1.2. implies the upper estimate in Theorem 1.1 by a sub-sequence extraction argument: if the upper estimate in Theorem 1.2 holds, then for all $n > 0$ there exists a sequence $(\mathcal{T}_N^n)_{N>N_0}$ such that

$$\limsup_{N\to+\infty}\left(N^{\frac{m}{2}} e_{\mathcal{T}_N^n}(f)\right) \le \left\|K\left(\frac{d^m f}{m!}\right)\right\|_{L^q} + \frac{1}{n},$$

with $\frac{1}{q} = \frac{1}{p} + \frac{m}{2}$. We then take $\mathcal{T}_N = \mathcal{T}_N^{n(N)}$, where

$$n(N) = \max\left\{ n \le N \; ; \; N^{\frac{m}{2}} e_{\mathcal{T}_N^n}(f) \le \left\| K\left(\frac{d^m f}{m!}\right)\right\|_{L^q} + \frac{2}{n} \right\}.$$

For N large enough this set is finite and non empty, and therefore $n(N)$ is well defined. Furthermore $n(N) \to +\infty$ as $N \to +\infty$ and therefore

$$\limsup_{N\to+\infty}\left(N^{\frac{m}{2}} e_{\mathcal{T}_N}(f)\right) \le \left\| K\left(\frac{d^m f}{m!}\right)\right\|_{L^q}.$$

We are thus left with proving the upper estimate in Theorem 1.2. We begin by fixing a (large) number $M > 0$. We shall take the limit $M \to \infty$ in the very last step of our proof. We define

$$\mathbb{T}_M = \{T \text{ triangle}, \, |T| = 1, \; \mathrm{bary}(T) = 0 \text{ and } \; \mathrm{diam}(T) \le M\},$$

the set of triangles centered at the origin, of unit area and diameter smaller than M. This set is compact with respect to the Hausdorff distance. This allows us to define a "tempered" version of $K = K_{m,p}$ that we denote by K_M:

$$K_M(\pi) = \inf_{T\in\mathbb{T}_M} e_T(\pi).$$

Since $\mathbb{T}_M$ is compact, the above infimum is attained on a triangle that we denote by $T_M(\pi)$. Note that the map $\pi \mapsto T_M(\pi)$ need not be continuous. It is clear that $K_M(\pi)$ decreases as M grows. Note also that the restriction to triangles T centered at 0 is artificial, since the error is invariant by translation as noticed in (8). Therefore $K_M(\pi)$ converges to $K(\pi)$ as $M \to +\infty$. Since $\mathbb{T}_M$ is compact, the map $\pi \mapsto \max_{T\in\mathbb{T}_M} e_T(\pi)$ defines a norm on $\mathbb{H}_m$, and is therefore bounded by $C_M\|\pi\|$ for some $C_M > 0$. One easily sees that the functions $\pi \mapsto e_T(\pi)$ are uniformly C_M-Lipschitz for all $T \in \mathbb{T}_M$, and so is K_M.

We now use this new function K_M to obtain a local upper error estimate that is closely related to the local lower estimate in Proposition 3.1

Proposition 3.2 *For $z_1 \in \Omega$, let T be a triangle which is obtained from $T_M(\pi_{z_1})$ by rescaling and translation ($T = tT_M(\pi_{z_1}) + z_0$). Then for any $z_2 \in T$,*

$$e_T(f) \le \Big(K_M(\pi_{z_2}) + B_M\omega(\max\{|z_1 - z_2|, \mathrm{diam}(T)\})\Big)|T|^{\frac{m}{2}+\frac{1}{p}}, \tag{22}$$

where $B_M > 0$ is a constant which depends on M.

Proof: For all $z_1, z_2 \in \Omega$, we have

$$\begin{aligned} e_{T_M(\pi_{z_1})}(\pi_{z_2}) &\le e_{T_M(\pi_{z_1})}(\pi_{z_1}) + C_M\|\pi_{z_1} - \pi_{z_2}\| \\ &= K_M(\pi_{z_1}) + C_M\|\pi_{z_1} - \pi_{z_2}\|, \\ &\le K_M(\pi_{z_2}) + 2C_M\|\pi_{z_1} - \pi_{z_2}\|, \\ &\le K_M(\pi_{z_2}) + 2C_M\omega(|z_1 - z_2|). \end{aligned}$$

Therefore, if T is of the form $T = tT_M(\pi_{z_1}) + z_0$, we obtain by a change of variable that

$$e_T(\pi_{z_2}) \le \Big(K_M(\pi_{z_2}) + 2C_M\omega(|z_1 - z_2|)\Big)|T|^{\frac{m}{2}+\frac{1}{p}}$$

Let $\mu_z \in \mathbb{P}_m$ be the Taylor polynomial of f at the point z up to degree m. Using Equation (17) we obtain

$$\begin{aligned} e_T(f) &\le e_T(\mu_{z_2}) + e_T(f - \mu_{z_2}) \\ &= e_T(\pi_{z_2}) + e_T(f - \mu_{z_2}) \\ &\le \Big(K_M(\pi_{z_2}) + 2C_M\omega(|z_1 - z_2|)\Big)|T|^{\frac{m}{2}+\frac{1}{p}} + e_T(f - \mu_{z_2}) \end{aligned}$$

By the same argument as in the proof of Proposition 3.1, we derive that

$$e_T(f-\mu_{z_2}) \le C|T|^{\frac{1}{p}}\,\mathrm{diam}(T)^m\omega(\mathrm{diam}\,T),$$

and thus

$$e_T(f) \le \Big(K_M(\pi_{z_2}) + 2C_M\omega(|z_1-z_2|)\Big)|T|^{\frac{m}{2}+\frac{1}{p}} + C|T|^{\frac{1}{p}}\,\mathrm{diam}(T)^m\omega(\mathrm{diam}\,T).$$

Since T is the scaled version of a triangle in $\mathbb{T}_M$, it obeys $\mathrm{diam}(T)^2 \le M^2|T|$. Therefore

$$e_T(f) \le (K_M(\pi_{z_2}) + (2C_M + CM^m)\omega(\max\{|z_1-z_2|, \mathrm{diam}(T)\}))\,|T|^{\frac{m}{2}+\frac{1}{p}},$$

which is the desired inequality with $B_M := 2C_M + CM^m$. $\diamond$

For some $r>0$ to be specified later, we now choose an arbitrary triangular mesh $\mathcal{R}$ of Ω satisfying

$$r \ge \sup_{R\in\mathcal{R}} \mathrm{diam}(R).$$

Our strategy to build a triangulation that satisfies the optimal upper estimate is to use the triangles R as *macro-elements* in the sense that each of them will be tiled by a locally optimal uniform triangulation. This strategy was already used in [4].

For all $R \in \mathcal{R}$ we consider the triangle

$$T_R := (K_M(\pi_{b_R}) + 2B_M\omega(r))^{-\frac{q}{2}}T_M(\pi_{b_R}),$$

which is a scaled version of $T_M(\pi_{b_R})$ where b_R is the barycenter of R. We use this triangle to build a periodic tiling $\mathcal{P}_R$ of the plane: there exists a vector c such that $T_R \cup T'_R$ forms a parallelogram of side vectors a and b, with $T'_R = c - T_R$. We then define

$$\mathcal{P}_R := \{T_R + ma + nb \ :\ m,n \in \mathbb{Z}^2\} \cup \{T'_R + ma + nb \ :\ m,n \in \mathbb{Z}^2\}. \tag{23}$$

Observe that for all $\pi \in \mathbb{H}_m$, and all triangles T, T' such that $T' = -T$ one has $e_T(\pi) = e_{T'}(\pi)$ since π is either an even polynomial when m is an even integer, or an odd polynomial when m is odd. Since we already know that $e_T(\pi)$ is invariant by translation of T, we find that the local error $e_T(\pi)$ is constant on all $T \in \mathcal{P}_R$.

We now define as follows a family of triangulations $\mathcal{T}_s$ of the domain Ω, for $s>0$. For every $R \in \mathcal{R}$, we consider the elements $T \cap R$ for $T \in s\mathcal{P}_R$, where $s\mathcal{P}_R$ denotes the triangulation $\mathcal{P}_R$ scaled by the factor s. Clearly $\{T\cap R,\ T \in s\mathcal{P}_R,\ R\in\mathcal{R}\}$ constitute a partition of Ω. In this partition, we distinguish the interior elements

$$\mathcal{T}_s^{\mathrm{reg}} := \{T \in s\mathcal{P}_R \ :\ T \in \mathrm{int}(R)\ , R \in \mathcal{R}\},$$

which define pieces of a conforming triangulation, and the boundary elements $T \cap R$ for $T \in s\mathcal{P}_R$ such that $T \cap \partial R \neq \emptyset$. These last elements might not be triangular, nor conformal with the elements on the other side. Note that for $s>0$ small enough, each $R \in \mathcal{R}$ contains at least one triangle in $\mathcal{T}_s^{\mathrm{reg}}$, and therefore the boundary elements constitute a layer around the edges of $\mathcal{R}$. In order to obtain a conforming triangulation, we proceed as follow: for each boundary element $T\cap R$, we consider the points on its boundary which are either its vertices or those of a neighboring element. We then build the Delaunay triangulation of these points, which is a triangulation of $T \cap R$ since it is a convex set. We denote by $\mathcal{T}_s^{\mathrm{bd}}$ the set of all triangles obtained by this procedure, which is illustrated on Figure 1.

Our conforming triangulation is given by

$$\mathcal{T}_s = \mathcal{T}_s^{\mathrm{reg}} \cup \mathcal{T}_s^{\mathrm{bd}}.$$

As $s \to 0$, clearly

$$\#(\mathcal{T}_s^{\mathrm{bd}}) \le C_{\mathrm{bd}}s^{-1} \text{ and } \sum_{T\in\mathcal{T}_s^{\mathrm{bd}}} |T| \le C_{\mathrm{bd}}s,$$

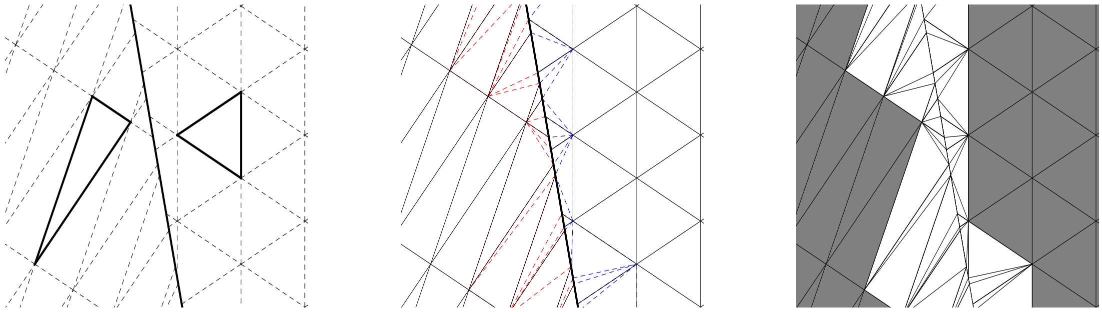

Figure 1: a. An edge (Thick) of the macro-triangulation $\mathcal{R}$ separating to uniformly paved regions (T_R is thick, $\mathcal{P}_R$ is dashed). b. Additional edges (dashed) are added near the interface in order to preserve conformity. c. The sets of triangles $\mathcal{T}_s^{\mathrm{reg}}$ (gray) and $\mathcal{T}_s^{\mathrm{bd}}$ (white)

for some constant C_{bd} which depends on the macro-triangulation $\mathcal{R}$. We do not need to estimate C_{bd} since $\mathcal{R}$ is fixed and the contribution due to C_{bd} in the following estimates is neglectible as $s \to 0$. We therefore obtain that the number of triangles in $\mathcal{T}_s^{\mathrm{bd}}$ is dominated by the number of triangles in $\mathcal{T}_s^{\mathrm{reg}}$. More precisely, we have the equivalence

$$\#(\mathcal{T}_s) \sim \#(\mathcal{T}_s^{\mathrm{reg}}) \sim \sum_{R\in\mathcal{R}} \frac{|R|}{s^2|T_R|} = s^{-2}\sum_{R\in\mathcal{R}} |R|(K_M(\pi_{b_R}) + 2B_M\omega(r))^q, \tag{24}$$

in the sense that the ratio between the above quantities tends to 1 as $s \to 0$. The right hand side in (24) can be estimated through an integral:

$$\begin{aligned} s^2\#(\mathcal{T}_s^{\mathrm{reg}}) &\le \sum_{R\in\mathcal{R}} |R|(K_M(\pi_{b_R}) + 2B_M\omega(r))^q \\ &= \sum_{R\in\mathcal{R}} \int_R (K_M(\pi_{b_R}) + 2B_M\omega(r))^q dz \\ &\le \sum_{R\in\mathcal{R}} \int_R (K_M(\pi_z) + C_M\|\pi_z - \pi_{b_R}\| + 2B_M\omega(r))^q dz \\ &\le \int_\Omega (K_M(\pi_z) + (2B_M + C_M)\omega(r))^q dz \end{aligned}$$

Therefore, since $C_M \le B_M$,

$$\#(\mathcal{T}_s) \le s^{-2}\left(\int_\Omega (K_M(\pi_z) + 3B_M\omega(r))^q dz + C_{\mathrm{bd}} s\right). \tag{25}$$

Observe that the construction of $\mathcal{T}_s$ gives a bound on the diameter of its elements

$$\sup_{T\in\mathcal{T}_s} \mathrm{diam}(T) \le sC_a, \quad C_a := \max_{R\in\mathcal{R}} \mathrm{diam}(T_R)$$

Combining this with (24), we obtain that

$$\sup_{T\in\mathcal{T}_s} \mathrm{diam}(T) \le C_A(\#\mathcal{T}_s)^{-1/2} \text{ for all } s > 0$$

which is analogous to the admissibility condition (5).

We now estimate the global interpolation error $\|f - I_{m,\mathcal{T}_s} f\|_{L^p} := (\sum_{T\in\mathcal{T}_s} e_T(f)^p)^{\frac{1}{p}}$, assuming first that $1 \le p < \infty$. We first estimate the contribution of $\mathcal{T}_s^{\mathrm{bd}}$, which will eventually be neglectible. Denoting $\nu_z \in \mathbb{P}_{m-1}$ the Taylor polynomial of f up to degree $m-1$ at z we remark that

$$\|f - I_{m,T}f\|_{L^\infty(T)} = \|(I - I_{m,T})(f - \nu_{b_T})\|_{L^\infty(T)} \le C_1\|f - \nu_{b_T}\|_{L^\infty(T)} \le C_0C_1\,\mathrm{diam}(T)^m.$$

where C_1 is the norm of $I - I_{m,T}$ in $L^\infty(T)$ which is independent of T and C_0 only depends on the L^∞ norm of $d^m f$. Remarking that $e_T(f) = \|f - I_{m,T} f\|_{L^p(T)} \le |T|^{\frac{1}{p}} \|f - I_{m,T} f\|_{L^\infty(T)}$, we obtain an upper bound for the contribution of $\mathcal{T}_s^{\rm bd}$ to the error:

$$\begin{aligned}
\sum_{T\in\mathcal{T}_s^{\rm bd}} e_T(f)^p &\le C_0^p C_1^p \sum_{T\in\mathcal{T}_s^{\rm bd}} |T| \operatorname{diam}(T)^{mp} \\
&\le C_0^p C_1^p \left(\sum_{T\in\mathcal{T}_s^{\rm bd}} |T|\right) \sup_{T\in\mathcal{T}_s^{\rm bd}} \operatorname{diam}(T)^{mp} \\
&\le C_0^p C_1^p C_{\rm bd} s \sup_{T\in\mathcal{T}_s^{\rm bd}} \operatorname{diam}(T)^{mp} \\
&\le C_{\rm bd}^* s^{mp+1},
\end{aligned}$$

with $C_{\rm bd}^* = C_0^p C_1^p C_a^{mp} C_{\rm bd}$. We next turn to the the contribution of $\mathcal{T}_s^{\rm reg}$ to the error. If $T \in \mathcal{T}_s^{\rm reg}$, $T \subset R \in \mathcal{R}$, we consider any point $z_1 = z \in T$ and define $z_2 = b_R$ the barycenter of R. With such choices, the estimate (22) reads

$$e_T(f) \le (K_M(\pi_z) + B_M \omega(\max\{r, C_A s\}))\, |T|^{\frac{m}{2}+\frac{1}{p}}.$$

We now assume that s is chosen small enough such that $C_A s \le r$. Geometrically, this condition ensures that the "micro-triangles" constituting $\mathcal{T}_s$ actually have a smaller diameter than the "macro-triangles" constituting $\mathcal{R}$. This implies

$$e_T(f)^p \le (K_M(\pi_z) + B_M\omega(r))^p\, |T|^{\frac{mp}{2}+1} \tag{26}$$

Given a triangle $T \in \mathcal{T}_s^{\rm reg}$, $T \subset R \in \mathcal{R}$, and a point $z \in T$, one has

$$\begin{aligned}
|T| &= s^2 \left(K_M(\pi_{b_R}) + 2B_M\omega(r)\right)^{-q} \\
&\le s^2 \left(K_M(\pi_z) - C_M\|\pi_z - \pi_{b_R}\| + 2B_M\omega(r)\right)^{-q} \\
&\le s^2 \left(K_M(\pi_z) + (2B_M - C_M)\omega(r)\right)^{-q}.
\end{aligned}$$

Observing that $B_M \ge C_M$, and that $p - q\frac{mp}{2} = q$, we inject the above inequality in the estimate (26), which yields

$$e_T(f)^p \le s^{mp} \left(K_M(\pi_z) + B_M\omega(r)\right)^q |T|.$$

Averaging on $z \in T$, we obtain

$$e_T(f)^p \le s^{mp} \int_T \left(K_M(\pi_z) + B_M\omega(r)\right)^q dz.$$

Adding up contributions from all triangles in $\mathcal{T}_s$, we find

$$e_{\mathcal{T}_s}(f)^p = \sum_{T\in\mathcal{T}_s^{\rm reg}} e_T(f)^p + \sum_{T\in\mathcal{T}_s^{\rm bd}} e_T(f)^p \le s^{mp} \int_\Omega \left(K_M(\pi_z) + B_M\omega(r)\right)^q dz + C_{\rm bd}^* s^{mp+1}$$

Combining this with the estimate (25) we obtain,

$$e_{\mathcal{T}_s} \#(\mathcal{T}_s)^{m/2} \le \left(\int_\Omega \left(K_M(\pi_z) + B_M\omega(r)\right)^q dz + C_{bd}^* s\right)^{\frac{1}{p}} \left(\int_\Omega \left(K_M(\pi_z) + 3B_M\omega(r)\right)^q dz + C_{\rm bd} s\right)^{\frac{m}{2}}$$

and therefore, since $\frac{1}{q} = \frac{m}{2} + \frac{1}{p}$,

$$\limsup_{s\to 0}\Big(\#(\mathcal{T}_s)^{m/2} e_{\mathcal{T}_s}\Big) \le \left(\int_\Omega \left(K_M(\pi_z) + 3B_M\omega(r)\right)^q dz\right)^{\frac{1}{q}}.$$

It is now time to observe that for fixed M,

$$\lim_{r\to 0}\int_\Omega (K_M(\pi_z)+3B_M\omega(r))^q\,dz = \int_\Omega K_M^q(\pi_z)dz,$$

and that

$$\lim_{M\to+\infty}\int_\Omega K_M^q(\pi_z)dz = \int_\Omega K^q(\pi_z)dz.$$

Therefore, for all $\varepsilon>0$, we can choose M sufficiently large and r sufficiently small, such that

$$\limsup_{s\to 0}\Big(\#(\mathcal{T}_s)^{m/2}e_{\mathcal{T}_s}\Big) \le \Big(\int_\Omega K^q(\pi_z)dz\Big)^{\frac{1}{q}}+\varepsilon.$$

This gives us the announced statement of Theorem 1.2, by defining

$$s_N := \min\{s>0\ :\ \#(T_s)\le N\},$$

and by setting $\mathcal{T}_N=\mathcal{T}_{s_N}$.

The adaptation of the above proof in the case $p=\infty$ is not straightforward due to the fact that the contribution to the error of $\mathcal{T}_s^{\rm bd}$ is not anymore neglectible with respect to the contribution of $\mathcal{T}_s^{\rm reg}$. For this reason, one needs to modify the construction of $\mathcal{T}_s^{\rm bd}$. Here, we provide a simple construction but for which the resulting triangulation $\mathcal{T}_s$ is non-conforming, as we do not know how to produce a satisfying conforming triangulation.

More precisely, we define $\mathcal{T}_s^{\rm reg}$ in a similar way as for $p<\infty$, and add to the construction of $\mathcal{T}_s^{\rm bd}$ a post processing step in which each triangle is splitted in 4^j similar triangles according to the midpoint rule. Here we take for j the smallest integer which is larger than $-\frac{\log s}{4\log 2}$. With such an additional splitting, we thus have

$$\max_{T\in\mathcal{T}_s^{\rm bd}}\operatorname{diam}(T)\le s^{\frac14}\max_{R\in\mathcal{R}}\operatorname{diam}(sT_R)=C_a s^{1+\frac14}.$$

The contribution of $\mathcal{T}_s^{\rm bd}$ to the L^∞ interpolation error is bounded by

$$e_{\mathcal{T}_s^{\rm bd}}(f)\le C_0C_1\max_{T\in\mathcal{T}_s^{\rm bd}}\operatorname{diam}(T)^m\le C_{\rm bd}^* s^{\frac{5m}{4}},$$

with $C_{\rm bd}^*:=C_0C_1C_a^m$. We also have

$$\#(\mathcal{T}_s^{\rm bd})\le C_{\rm bd}s^{-3/2},$$

which remains neglectible compared to s^{-2}. We therefore obtain

$$\#(\mathcal{T}_s)\le s^{-2}\left(\int_\Omega (K_M(\pi_z)+3B_M\omega(r))^{\frac{2}{m}}dz + C_{\rm bd}s^{1/2}\right) \tag{27}$$

Moreover, if $T\in\mathcal{T}_s^{\rm reg}$ and $T\subset R\in\mathcal{R}$, we have according to the estimate (22)

$$e_T(f)\le (K_M(\pi_{b_R})+B_M\omega(\max\{r,C_As\}))\,|T|^{\frac{m}{2}}.$$

By construction $|T|=s^2(K_M(\pi_{b_R})+2B_M\omega(r))^{-2/m}$. This implies $e_T(f)\le s^m$ when $C_As\le r$. Therefore

$$e_{\mathcal{T}_s}(f)=\max\{e_{\mathcal{T}_s^{\rm reg}},e_{\mathcal{T}_s^{\rm bd}}\}\le s^m\max\{1,C_{\rm bd}^*s^{\frac{m}{4}}\}.$$

Combining this estimate with (27) yields

$$\limsup_{s\to 0}\Big(\#(\mathcal{T}_s)^{m/2}e_{\mathcal{T}_s}\Big)\le\left(\int_\Omega (K_M(\pi_z)+3B_M\omega(r))^{\frac{2}{m}}\,dz\right)^{\frac{m}{2}},$$

and we conclude the proof in a similar way as for $p<\infty$.

4 The shape function and the optimal metric for linear and quadratic elements

This section is devoted to linear ($m = 2$) and quadratic ($m = 3$) elements, which are the most commonly used in practice. In these two cases, we are able to derive an exact expression for $K_{m,p}(\pi)$ in terms of the coefficients of π. Our analysis also gives us access to the distorted metric which characterizes the optimal mesh. While the results concerning linear elements have strong similarities with those of [4], those concerning quadratic elements are to our knowledge the first of this kind, although [10] analyzes a similar setting.

4.1 Exact expression of the shape function

In order to give the exact expression of $K_{m,p}$, we define the determinant of an homogeneous quadratic polynomial by

$$\det(ax^2 + 2bxy + cy^2) = ac - b^2,$$

and the discriminant of an homogeneous cubic polynomial by

$$\operatorname{disc}(ax^3 + bx^2y + cxy^2 + dy^3) = b^2c^2 - 4ac^3 - 4b^3d + 18abcd - 27a^2d^2.$$

The functions det on $\mathbb{H}_2$ and disc on $\mathbb{H}_3$ are homogeneous in the sense that

$$\det(\lambda\pi) = \lambda^2 \det\pi, \quad \operatorname{disc}(\lambda\pi) = \lambda^4 \operatorname{disc}\pi. \tag{28}$$

Moreover, it is well known that they obey an invariance property with respect to linear changes of coordinates ϕ:

$$\det(\pi \circ \phi) = (\det\phi)^2 \det\pi, \quad \operatorname{disc}(\pi \circ \phi) = (\det\phi)^6 \operatorname{disc}\pi. \tag{29}$$

Our main result relates $K_{m,p}$ to these quantities.

Theorem 4.1 *We have for all $\pi \in \mathbb{H}_2$,*

$$K_{2,p}(\pi) = \sigma_p(\det\pi)\sqrt{|\det\pi|},$$

and for all $\pi \in \mathbb{H}_3$,

$$K_{3,p}(\pi) = \sigma_p^*(\operatorname{disc}\pi)\sqrt[4]{|\operatorname{disc}\pi|},$$

where $\sigma_p(t)$ and $\sigma_p^*(t)$ are constants that only depend on the sign of t.

The proof of Theorem 4.1 relies on the possibility of mapping and arbitrary polynomial $\pi \in \mathbb{H}_2$ such that $\det(\pi) \neq 0$ or $\pi \in \mathbb{H}_3$ such that $\operatorname{disc}(\pi) \neq 0$ onto two fixed polynomials π_- or π_+ by a linear change of variable and a sign change.

In the case of $\mathbb{H}_2$, it is well known that we can choose $\pi_- = x^2 - y^2$ and $\pi_+ = x^2 + y^2$. More precisely, to all $\pi \in H_2$, we associate a symmetric matrix Q_π such that $\pi(z) = \langle Q_\pi z, z\rangle$. This matrix can be diagonalized according to

$$Q_\pi = U^{\mathrm{T}} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} U, \quad U \in \mathcal{O}_2,\ \lambda_1, \lambda_2 \in \mathbb{R}.$$

Then, defining the linear transform

$$\phi_\pi := U^{\mathrm{T}} \begin{pmatrix} |\lambda_1|^{-\frac{1}{2}} & 0 \\ 0 & |\lambda_2|^{-\frac{1}{2}} \end{pmatrix}$$

and $\lambda_\pi = \operatorname{sign}(\lambda_1) \in \{-1, 1\}$, it is readily seen that

$$\lambda_\pi \pi \circ \phi_\pi = \begin{cases} x^2 + y^2 & \text{if } \det\pi > 0 \\ x^2 - y^2 & \text{if } \det\pi < 0. \end{cases}$$

In the case of $\mathbb{H}_3$, a similar result holds, as shown by the following lemma.

Lemma 4.1 *Let $\pi \in \mathbb{H}_3$. There exists a linear transform ϕ_π such that*

$$\pi \circ \phi_\pi = \begin{cases} x(x^2 - 3y^2) & \text{if } \operatorname{disc}\pi > 0 \\ x(x^2 + 3y^2) & \text{if } \operatorname{disc}\pi < 0. \end{cases} \tag{30}$$

Proof: Let us first assume that π is not divisible by y so that it can be factorized as

$$\pi = \lambda(x - r_1 y)(x - r_2 y)(x - r_3 y),$$

with $\lambda \in \mathbb{R}$ and $r_i \in \mathbb{C}$. If $\operatorname{disc} \pi > 0$, then the r_i are real and we may assume $r_1 < r_2 < r_3$. Then, defining

$$\phi_\pi = \lambda(2 \operatorname{disc} \pi)^{-1/3} \begin{pmatrix} r_1(r_2 + r_3) - 2r_2 r_3 & (r_2 - r_3) r_1 \sqrt{3} \\ 2r_1 - (r_2 + r_3) & (r_2 - r_3)\sqrt{3}, \end{pmatrix}.$$

an elementary computation shows that $\pi \circ \phi_\pi = x(x^2 - 3y^2)$. If $\operatorname{disc} \pi < 0$, then we may assume that r_1 is real, r_2 and r_3 are complex conjugates with $\operatorname{Im}(r_2) > 0$. Then, defining

$$\phi_\pi = \lambda(2 \operatorname{disc} \pi)^{-1/3} \begin{pmatrix} r_1(r_2 + r_3) - 2r_2 r_3 & \mathbf{i}(r_2 - r_3) r_1 \sqrt{3} \\ 2r_1 - (r_2 + r_3) & \mathbf{i}(r_2 - r_3)\sqrt{3} \end{pmatrix}.$$

an elementary computation shows that $\pi \circ \phi_\pi = x(x^2 + 3y^2)$. Moreover it is easily checked that ϕ_π has real entries and is therefore a change of variable in $\mathbb{R}^2$.

In the case where π is divisible by y, there exists a rotation $U \in \mathcal{O}_2$ such that $\tilde{\pi} := \pi \circ U$ is not divisible by y. By the invariance property (29) we know that $\operatorname{disc} \pi = \operatorname{disc} \tilde{\pi}$. Thus, we reach the same conclusion with the choice $\phi_\pi := U \circ \phi_{\tilde{\pi}}$. $\diamond$

Proof of Theorem 4.1: for all $\pi \in \mathbb{H}_2$ such that $\det \pi \neq 0$ and for all change of variable ϕ and $\lambda \neq 0$, we may combine the properties of the determinant in (28) and (29) with those of the shape function established in Proposition 2.2. This gives us

$$\frac{K_{2,p}(\pi)}{\sqrt{|\det \pi|}} = \frac{K_{2,p}(\lambda \pi \circ \phi)}{\sqrt{|\det(\lambda \pi \circ \phi)|}}.$$

Applying this with $\phi = \phi_\pi$ and $\lambda = \lambda_\pi$, we therefore obtain

$$K_{2,p}(\pi) = \sqrt{|\det \pi|} \begin{cases} K_{2,p}(x^2 + y^2) & \text{if } \det \pi > 0, \\ K_{2,p}(x^2 - y^2) & \text{if } \det \pi < 0. \end{cases}$$

This gives the desired result with $\sigma_p(t) = K_{2,p}(x^2 + y^2)$ for $t > 0$ and $\sigma_p(t) = K_{2,p}(x^2 - y^2)$ for $t < 0$. In the case where $\det \pi = 0$, then π is of the form $\pi(x, y) = \lambda(\alpha x + \beta y)^2$ and we conclude by Proposition 2.1 that $K_{2,p}(\pi) = 0$.

For all $\pi \in \mathbb{H}_3$ such that $\operatorname{disc} \pi \neq 0$, a similar reasoning yields

$$K_{3,p}(\pi) = \sqrt[4]{|\operatorname{disc} \pi|}\, 108^{-\frac{1}{4}} \begin{cases} K_{3,p}(x(x^2 - 3y^2)) & \text{if } \operatorname{disc} \pi > 0, \\ K_{3,p}(x(x^2 + 3y^2)) & \text{if } \operatorname{disc} \pi < 0. \end{cases},$$

where the constant 108 comes from the fact that $\operatorname{disc}(x(x^2 - 3y^2)) = -\operatorname{disc}(x(x^2 - 3y^2)) = 108$. This gives the desired result with $\sigma_p^*(t) = 108^{-\frac{1}{4}} K_{3,p}(x(x^2 - 3y^2))$ for $t > 0$ and $\sigma_p^*(t) = 108^{-\frac{1}{4}} K_{3,p}(x(x^2 + 3y^2))$ for $t < 0$. In the case where $\operatorname{disc} \pi = 0$, then π is of the form $\pi(x, y) = (\alpha x + \beta y)^2(\gamma x + \delta y)$ and we conclude by Proposition 2.1 that $K_{3,p}(\pi) = 0$. $\diamond$

Remark 4.2 *We do not know any simple analytical expression for the constants involved in σ_p and σ_p^*, but these can be found by numerical optimization. These constants are known for some special values of p in the case $m = 2$, see for example [4].*

4.2 Optimal metrics

Practical mesh generation techniques such as in [20, 6, 7, 26, 27] are based on the data of a Riemannian metric, by which we mean a field h of symmetric definite positive matrices

$$x \in \Omega \mapsto h(x) \in S_2^+.$$

Typically, the mesh generator takes the metric h as an input and hopefully returns a triangulation $\mathcal{T}_h$ adapted to it in the sense that all triangles are close to equilateral of unit side length with respect to this metric. Recently, it has been rigorously proved in [24, 6] that some algorithms produce bidimensional meshes obeying these constraints, under certain conditions. This must be contrasted with algorithms based on heuristics, such as [26] in two dimensions, and [27] in three dimensions, which have been available for some time and offer good performance [8] but no theoretical guaranties.

For a given function f to be approximated, the field of metrics given as input should be such that the local errors are equidistributed and the aspect ratios are optimal for the generated triangulation. Assuming that the error is measured in $X = L^p$ and that we are using finite elements of degree $m-1$, we can construct this metric as follows, provided that some estimate of $\pi_z = \frac{d^m f(z)}{m!}$ is available all points $z \in \Omega$. An ellipse E_z such that $|E_z|$ is equal or close to

$$\sup_{E\in\mathcal{E}, E\subset\Lambda_{\pi_z}} |E| \tag{31}$$

is computed, where Λ_{π_z} is defined as in (13). We denote by $h_{\pi_z} \in S_2^+$ the associated symmetric definite positive matrix such that

$$E_z = \left\{(x,y) \;:\; (x,y)^T h_{\pi_z}(x,y) \le 1\right\}.$$

Let us notice that the supremum in (31) might not always be attained or even be finite. This particular case is discussed in the end of this section. Denoting by $\nu > 0$ the desired order of the L^p error on each triangle, we then define the metric by rescaling h_{π_z} according to

$$h(z) = \frac{1}{\alpha_z^2} h_{\pi_z} \;\text{ where }\; \alpha_z := \nu^{\frac{p}{mp+2}} |E_z|^{-\frac{1}{mp+2}}.$$

With such a rescaling, any triangle T designed by the mesh generator should be comparable to the ellipse $z + \alpha_z E_z$ centered around z the barycenter of T, in the sense that

$$z + c_1\alpha_z E_z \subset T \subset z + c_2\alpha_z E_z, \tag{32}$$

for two fixed constants $0 < 2c_1 \le c_2$ independent of T (recall that for any ellipse E there always exist a triangle T such that $E \subset T \subset 2E$).

Such a triangulation heuristically fulfills the desired properties of optimal aspect ratio and error equidistribution when the level of refinement is sufficiently small. Indeed, we then have

$$\begin{array}{rcl}
e_{m,T}(f)_p & \approx & e_{m,T}(\pi_z)_p \\
& = & \|\pi_z - I_{m,T}\pi_z\|_{L^p(T)}, \\
& \sim & |T|^{\frac{1}{p}} \|\pi_z - I_{m,T}\pi_z\|_{L^\infty(T)}, \\
& \sim & |T|^{\frac{1}{p}} \|\pi_z\|_{L^\infty(T)}, \\
& \sim & |\alpha_z E_z|^{\frac{1}{p}} \|\pi_z\|_{L^\infty(\alpha_z E_z)}, \\
& = & \alpha_z^{m+\frac{2}{p}} |E_z|^{\frac{1}{p}} \|\pi_z\|_{L^\infty(E_z)}, \\
& = & \nu,
\end{array}$$

where we have used the fact that $\pi_z \in \mathbb{H}_m$.

Leaving aside these heuristics on error estimation and mesh generation, we focus on the main computational issue in the design of the metric $h(z)$, namely the solution to the problem (31): to any given $\pi \in \mathbb{H}_m$, we want to associate $h_\pi \in S_2^+$ such that the ellipse E_π defined by h_π has area equal or close to $\sup_{E\in\mathcal{E}, E\subset\Lambda_\pi} |E|$.

When $m = 2$ the computation of the optimal matrix h_π can be done by elementary algebraic means. In fact, as it will be recalled below, h_π is simply the absolute value (in the sense of symmetric matrices) of the symmetric matrix $[\pi]$ associated to the quadratic form π. These facts are well known and used in mesh generation algorithms for $\mathbb{P}_1$ elements.

When $m \ge 3$ no such algebraic derivation of h_π from π has been proposed up to now and current approaches instead consist in numerically solving the optimization problem (15), see [9]. Since these

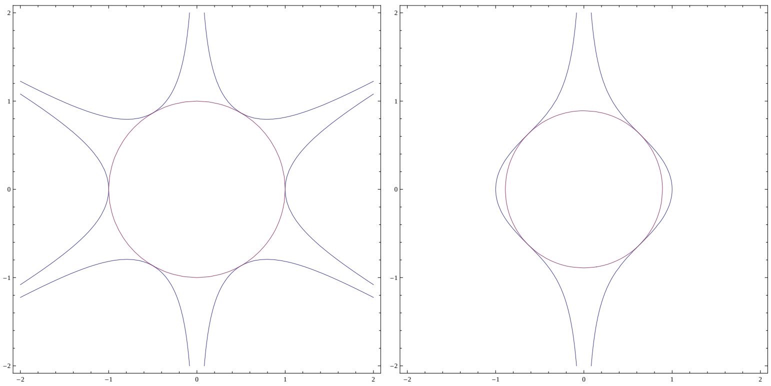

Figure 2: Maximal ellipses inscribed in Λ_π, $\pi = x(x^2-3y^2)$ or $\pi = x(x^2+3y^2)$.

computations have to be done extremely frequently in the mesh adaptation process, a simpler algebraic procedure is highly valuable. In this section, we propose a simple and algebraic method in the case $m=3$, corresponding to quadratic elements. For purposes of comparison the results already known in the case $m=2$ are recalled.

Proposition 4.2 *1. Let $\pi \in \mathbb{H}_2$ be such that $\det(\pi) \neq 0$, and consider its associated 2×2 matrix which can be written as*

$$[\pi] = U^{\mathrm{T}} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} U, \quad U \in \mathcal{O}_2.$$

Then, an ellipse of maximal volume inscribed in Λ_π is defined by the matrix

$$h_\pi = U^{\mathrm{T}} \begin{pmatrix} |\lambda_1| & 0 \\ 0 & |\lambda_2| \end{pmatrix} U$$

2. Let $\pi \in \mathbb{H}_3$ be such that $\operatorname{disc}\pi > 0$, and ϕ_π a matrix satisfying (30). Define

$$h_\pi = (\phi_\pi^{-1})^{\mathrm{T}} \phi_\pi^{-1}. \tag{33}$$

Then h_π defines an ellipse of maximal volume inscribed in Λ_π. Moreover $\det h_\pi = \frac{2^{-2/3}}{3} (\operatorname{disc}\pi)^{\frac{1}{3}}$.

3. Let $\pi \in \mathbb{H}_3$ be such that $\operatorname{disc}\pi < 0$, and ϕ_π a matrix satisfying (30). Define

$$h_\pi = 2^{\frac{1}{3}} (\phi_\pi^{-1})^{\mathrm{T}} \phi_\pi^{-1}.$$

Then h_π defines an ellipse of maximal volume inscribed in Λ_π. Moreover $\det h_\pi = \frac{1}{3} |\operatorname{disc}\pi|^{\frac{1}{3}}$.

Proof: Clearly, if the matrix h_π defines an ellipse of maximal volume in the set Λ_π, then for any linear change of coordinates ϕ, the metric $(\phi^{-1})^{\mathrm{T}} h_\pi \phi^{-1}$ defines an ellipse of maximal volume in the set $\Lambda_{\pi\circ\phi}$. When $\pi \in \mathbb{H}_2$, we know that $\lambda_\pi \pi \circ \phi_\pi = x^2+y^2$ when $\det\pi > 0$, and x^2-y^2 when $\det\pi < 0$, where $|\lambda_\pi| = 1$. When $\pi \in \mathbb{H}_3$, we know from Lemma 4.1 that $\pi \circ \phi_\pi = x(x^2-3y^2)$ when $\operatorname{disc}\pi > 0$ and $x(x^2+3y^2)$ when $\operatorname{disc}\pi < 0$. Hence it only remains to prove that when $\pi \in \{x^2+y^2,\ x^2-y^2,\ x(x^2-3y^2)\}$, then $h_\pi = \mathrm{Id}$, which means that the disc of radius 1 is an ellipse of maximal volume inscribed in Λ_π, while when $\pi = x(x^2+3y^2)$ we have $h_\pi = 2^{1/3}\,\mathrm{Id}$.

The case $\pi = x^2+y^2$ is trivial. We next concentrate on the case $\pi = x(x^2+3y^2)$, the treatment of the two other cases being very similar. Let E be an ellipse included in Λ_π, $\pi = x(x^2+3y^2)$. Analyzing the variations of the function $\pi(\cos\theta, \sin\theta)$, it is not hard to see that we can rotate E into another ellipse E', also verifying the inclusion $E' \subset \Lambda_\pi$, and which principal axes are $\{x=0\}$ and $\{y=0\}$. We therefore only need to consider ellipses of the form $kx^2+hy^2 \leq 1$. For a given value of h, we denote by $k(h)$ the minimal value of k for which this ellipse is included in Λ_π. Clearly the boundary of the ellipse, defined by $k(h)x^2 + hy^2 = 1$, must be tangent to the curve defined by $\pi(x, y) = 1$ at some point (x, y). This translates into the following system of equations

$$\left\{ \begin{array}{rcl} \pi(x,y) & = & 1, \\ hx^2 + ky^2 & = & 1, \\ ky\partial_x\pi(x,y) - hx\partial_x\pi(x,y) & = & 0. \end{array} \right. \tag{34}$$

Eliminating the variables x and y from this system, as well as negative or complex valued solutions, we find that $k(h) = \frac{4+h^3}{3h^2}$ when $h \in (0, 2]$, and $k(h) = k(2) = 1$ when $h \geq 2$. The minimum of the determinant $hk(h) = \frac{1}{3}\left(\frac{4}{h} + h^2\right)$ is attained for $h = 2^{\frac{1}{3}}$. Observing that $k(2^{\frac{1}{3}}) = 2^{\frac{1}{3}}$ we obtain as announced $h_\pi = 2^{1/3}\,\mathrm{Id}$ and that the ellipse of largest area included in Λ_π is the disc of equation $2^{1/3}(x^2 + y^2) \leq 1$, as illustrated on Figure 2.b.

The same reasoning applies to the other cases. For $\pi = x^2 - y^2$ we obtain $k(h) = \frac{1}{h}$, $h \in (0, \infty)$. In this case the determinant $hk(h)$ is independent of h, and we simply choose $h = 1 = k(1)$. For $\pi = x(x^2 - 3y^2)$ we obtain $k(h) = \frac{4-h^3}{3h^2}$ when $h \in (0, 1]$ and $k(h) = k(1) = 1$ when $h > 1$. The maximal volume is attained when $h = 1$, corresponding to the unit disc, as illustrated on Figure 2.a. $\diamond$

Remark 4.3 *When $\pi \in \mathbb{H}_3$ and* $\operatorname{disc}\pi > 0$ *a surprising simplification happens : the matrix* (33) *has entries which are symmetric functions of the roots r_1, r_2, r_3. Using the relation between the roots and the coefficients of a polynomial, we find the following expression*

$$\text{If } \pi = ax^3 + 3bx^2y + 3cxy^2 + dy^3, \text{ then } h_\pi = 2^{-\frac{1}{3}}3(\operatorname{disc}\pi)^{\frac{-1}{3}} \begin{pmatrix} 2(b^2 - ac) & bc - ad \\ bc - ad & 2(c^2 - bd) \end{pmatrix}.$$

This yields a direct expression of the matrix as a function of the coefficients. Unfortunately there is no such expression when $\operatorname{disc}\pi < 0$.

At first sight, Proposition 4.2 might seem to be a complete solution to the problem of building an appropriate metric for mesh generation. However, some difficulties arise at points $z \in \Omega$ where $\det \pi_z = 0$ or $\operatorname{disc}\pi_z = 0$. If $\pi \in \mathbb{H}_2\backslash\{0\}$ and $\det \pi = 0$, then up to a linear change of coordinates, and a change of sign, we can assume that $\pi = x^2$. The minimization problem clearly yields the degenerate matrix $h_\pi = \operatorname{diag}(1, 0)$, the 2×2 diagonal matrix with entries 1 and 0. If $\pi \in \mathbb{H}_3\backslash\{0\}$ and $\operatorname{disc}\pi = 0$, then up to a linear change of coordinates either $\pi = x^3$ or $\pi = x^2y$. In the first case the minimization problem gives again $h_\pi = \operatorname{diag}(1, 0)$. In the second case a wilder behavior appears, in the sense that minimizing sequences for the problem (31) are of the type $h_\pi = \operatorname{diag}(\varepsilon^{-1}, \varepsilon^2)$ with $\varepsilon \to 0$. The minimization process therefore gives a matrix which is not only degenerate, but also unbounded.

These degenerate cases appear generically, and constitute a problem for mesh generation since they mean that the adapted triangles are not well defined. Current anisotropic mesh generation algorithms for linear elements often solve this problem by fixing a small parameter $\delta > 0$, and working with the modified matrix $\tilde{h}_\pi := h_\pi + \delta\,\mathrm{Id}$ which cannot degenerate. However this procedure cannot be extended to quadratic elements, since h_{x^2y} is both degenerate and unbounded.

In the theoretical construction of an optimal mesh which was discussed in §3.2, we tackled this problem by imposing a bound $M > 0$ on the diameter of the triangles. This was the purpose of the modified shape function $K_M(\pi)$ and of the triangle $T_M(\pi)$ of minimal interpolation error among the triangles of diameter smaller than M. We follow a similar idea here, looking for the ellipse of largest area included in Λ_π with constrained diameter. This provides matrices which are both positive definite and bounded, and vary continuously with respect to the data $\pi \in \mathbb{H}_3$. The constrained problem, depending on $\alpha > 0$, is the following:

$$\sup\{|E| \ : \ E \in \mathcal{E},\ E \subset \Lambda_\pi \text{ and } \operatorname{diam} E \leq 2\alpha^{-1/2}\}, \tag{35}$$

or equivalently

$$\inf\{\det H \ : \ H \in S_2^+ \text{ s.t. } \langle Hz, z\rangle \geq |\pi(z)|^{2/m}, z \in \mathbb{R}^2, \text{ and } H \geq \alpha\,\mathrm{Id}\}. \tag{36}$$

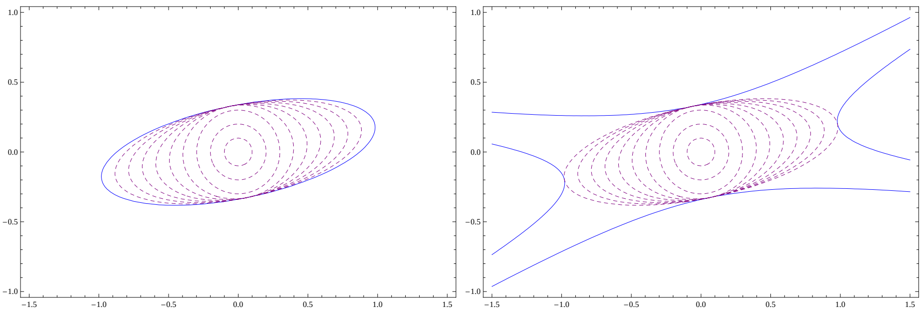

Figure 3: The set Λ_π (full) and the ellipses $E_{\pi,\alpha}$ (dashed) for various values of $\alpha > 0$ when $\pi \in \mathbb{H}_2$.

We denote by $E_{\pi,\alpha}$ and $h_{\pi,\alpha}$ the solutions to (35) and (36). In the remainder of this section, we show that this solution can also be computed by a simple algebraic procedure, avoiding any kind of numerical optimization. In the case where $\pi \in \mathbb{H}_2$, it can easily be checked that

$$[h_{\pi,\alpha}] = U^{\mathrm{T}} \begin{pmatrix} \max\{|\lambda_1|,\alpha\} & 0 \\ 0 & \max\{|\lambda_2|,\alpha\}, \end{pmatrix} U \tag{37}$$

as illustrated on Figure 3.

When $\pi \in \mathbb{H}_3$, the problem is more technical, and the matrix $h_{\pi,\alpha}$ takes different forms depending on the value of α and the sign of $\operatorname{disc}\pi$. In order to describe these different regimes, we introduce three real numbers $0 \le \beta_\pi \le \alpha_\pi \le \mu_\pi$ and a matrix $U_\pi \in \mathcal{O}_2$ which are defined as follow. We first define μ_π by

$$\mu_\pi^{-1/2} := \min\{\|z\| \ : \ |\pi(z)| = 1\},$$

the radius of the largest disc D_π inscribed in Λ_π. For z_π such that $|\pi(z_\pi)| = 1$ and $\|z_\pi\| = \mu_\pi^{-1/2}$, we define U_π as the rotation which maps z_π to the vector $(\|z_\pi\|, 0)$. We then define α_π by

$$2\alpha_\pi^{-1/2} := \max\{\operatorname{diam}(E) \ : \ E \in \mathcal{E} \ ; \ D_\pi \subset E \subset \Lambda_\pi\},$$

the diameter of the largest ellipse inscribed in Λ_π and containing the disc D_π. In the case where π is of the form $(ax+by)^3$, this ellipse is infinitely long and we set $\alpha_\pi = 0$. We finally define β_π by

$$2\beta_\pi^{-1/2} := \operatorname{diam}(E_\pi),$$

where E_π is the optimal ellipse described in Proposition 4.2. In the case where $\operatorname{disc}\pi = 0$, the "optimal ellipse" is infinitely long and we set $\beta_\pi = 0$. It is readily seen that $0 \le \beta_\pi \le \alpha_\pi \le \mu_\pi$.

All these quantities can be algebraically computed from the coefficients of π by solving equations of degree at most 4, as well as the other quantities involved in the description of the optimal $h_{\pi,\alpha}$ and $E_{\pi,\alpha}$ in the following result.

Proposition 4.3 *For $\pi \in \mathbb{H}_3$ and $\alpha > 0$, the matrix $h_{\pi,\alpha}$ and ellipse $E_{\pi,\alpha}$ are described as follows.*

1. *If $\alpha \ge \mu_\pi$, then $h_{\pi,\alpha} = \alpha\,\mathrm{Id}$ and $E_{\alpha,\pi}$ is the disc of radius $\alpha^{-1/2}$.*

2. *If $\alpha_\pi \le \alpha \le \mu_\pi$, then*
$$h_{\pi,\alpha} = U_\pi^{\mathrm{T}} \begin{pmatrix} \mu_\pi & 0 \\ 0 & \alpha \end{pmatrix} U_\pi, \tag{38}$$
and E_α is the ellipse of diameter $2\alpha^{-1/2}$ which is inscribed in Λ_π and contains D_π. It is tangent to $\partial\Lambda_\pi$ at the two points z_π and $-z_\pi$.

3. *If $\beta_\pi \le \alpha \le \alpha_\pi$ then $E_{\pi,\alpha}$ is tangent to $\partial\Lambda_\pi$ at four points and has diameter $2\alpha^{-1/2}$. There are at most three such ellipses and $E_{\pi,\alpha}$ is the one of largest area. The matrix $h_{\pi,\alpha}$ has a form which*

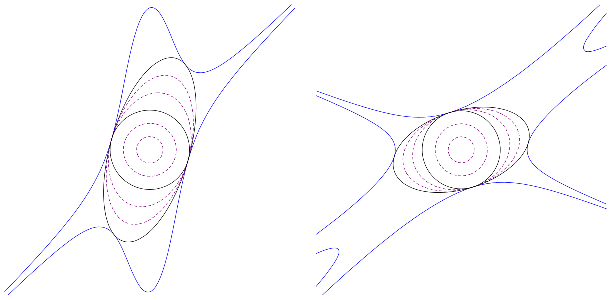

Figure 4: The set Λ_π (full), the disc $E_{\pi,\mu_\pi} = D_\pi$ (full), the ellipse E_{π,α_π} (full), and the ellipses $E_{\pi,\alpha}$ (dashed) for various values of $\alpha > 0$ when $\pi \in \mathbb{H}_3$ and $\alpha \in (\alpha_\pi, \infty)$. Left: $\operatorname{disc} \pi < 0$. Right: $\operatorname{disc} \pi > 0$

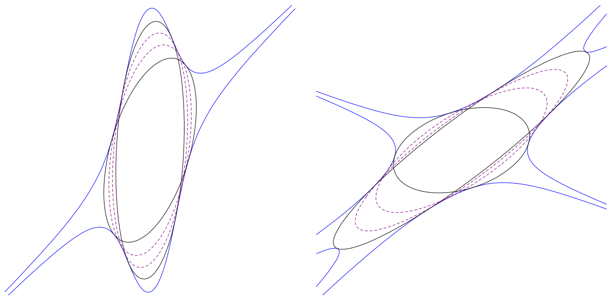

Figure 5: The set Λ_π (full), the ellipse E_{π,α_π} (full), the ellipse $E_{\pi,\beta_\pi} = E_\pi$ (full), and the ellipses $E_{\pi,\alpha}$ (dashed) for various values of $\alpha > 0$ when $\pi \in \mathbb{H}_3$ and $\alpha \in (\beta_\pi, \alpha_\pi)$. Left: $\operatorname{disc} \pi < 0$. Right: $\operatorname{disc} \pi > 0$

depends on the sign of $\operatorname{disc} \pi$.
(i) If $\operatorname{disc} \pi < 0$*, then*

$$h_{\pi,\alpha} = (\phi_\pi^{-1})^{\mathrm{T}} \begin{pmatrix} \lambda_\alpha & 0 \\ 0 & \frac{4+\lambda_\alpha^3}{3\lambda_\alpha^2} \end{pmatrix} \phi_\pi^{-1}$$

where ϕ_π *is the matrix defined in Proposition* 4.2 *and* λ_α *determined by* $\det(h_{\pi,\alpha} - \alpha \,\mathrm{Id}) = 0$.
(ii) If $\operatorname{disc} \pi > 0$*, then*

$$h_{\pi,\alpha} = (\phi_\pi^{-1})^{\mathrm{T}} V^{\mathrm{T}} \begin{pmatrix} \lambda_\alpha & 0 \\ 0 & \frac{4-\lambda_\alpha^3}{3\lambda_\alpha^2} \end{pmatrix} V \phi_\pi^{-1}$$

where ϕ_π *and* λ_α *are given as in the case* $\operatorname{disc} \pi < 0$ *and where* V *is chosen between the three rotations by* 0*,* 60 *or* 120 *degrees so to maximize* $|E_{\alpha,\pi}|$.
(iii) If $\operatorname{disc} \pi = 0$ *and* $\alpha_\pi > 0$*, then there exists a linear change of coordinates* ϕ *such that* $\pi \circ \phi = x^2 y$ *and we have*

$$h_{\pi,\alpha} = (\phi^{-1})^{\mathrm{T}} \begin{pmatrix} \lambda_\alpha & 0 \\ 0 & \frac{4}{27\lambda_\alpha^2} \end{pmatrix} \phi^{-1}$$

where λ_α *is determined by* $\det(h_{\pi,\alpha} - \alpha \,\mathrm{Id}) = 0$.

4. *If* $\alpha \leq \beta_\pi$*, then* $h_{\pi,\alpha} = h_\pi$ *and* $E_{\pi,\alpha} = E_\pi$ *is the solution of the unconstrained problem.*

Proof: See Appendix. ◇

Figure 4 illustrates the ellipses $E_{\pi,\alpha}$, $\alpha \in (\alpha_\pi, \infty)$ when $\operatorname{disc} \pi > 0$ (4.a) or $\operatorname{disc} \pi < 0$ (4.b). Figure 5 illustrates the ellipses $E_{\pi,\alpha}$, $\alpha \in (\beta_\pi, \alpha_\pi)$ when $\operatorname{disc} \pi > 0$ (5.a) or $\operatorname{disc} \pi < 0$ (5.b). Note that when $\alpha \geq \alpha_\pi$, the principal axes of $E_{\pi,\alpha}$ are independent of α since U_π is a rotation that only depends on π, while these axes generally vary when $\beta_\pi \leq \alpha \leq \alpha_\pi$, since the matrix ϕ_π is not a rotation.

Remark 4.4 *For interpolation by cubic or higher degree polynomials ($m \geq 4$), an additional difficulty arises that can be summarized as follows: one should be careful not to "overfit" the polynomial π with the matrix h_π. An approach based on exactly solving the optimization problem (31) might indeed lead to a metric $h(z)$ with unjustified strong variations with respect to z and/or bad conditioning, and jeopardize the mesh generation process. As an example, consider the one parameter family of polynomials*

$$\pi_t = x^2y^2 + ty^4 \in I\!H_4, \quad t \in [-1, 1].$$

It can be checked that when $t > 0$, the supremum $S_+ = \sup_{E \in \mathcal{E}, E \subset \Lambda_{\pi_t}} |E|$ is finite and independent of t, but not attained, and that any sequence $E_n \subset \Lambda_{\pi_t}$ of ellipses such that $\lim_{n\to\infty} |E_n| = S_+$ becomes infinitely elongated in the x direction, as $n \to \infty$. For $t < 0$, the supremum $S_- = \sup_{E \in \mathcal{E}, E \subset \Lambda_{\pi_t}} |E|$ is independent of t and attained for the optimal ellipse of equation $|t|^{-1/2}\frac{\sqrt{2}-1}{2}x^2 + |t|^{1/2}y^2 \leq 1$. This ellipse becomes infinitely elongated in the y direction as $t \to 0$. This example shows the instability of the optimal matrix h_π with respect to small perturbations of π. However, for all values of $t \in [-1, 1]$, these extremely elongated ellipses could be discarded in favor, for example, of the unit disc $D = \{x^2 + y^2 \leq 1\}$ which obviously satisfies $D \subset \Lambda_{\pi_t}$ and is a near-optimal choice in the sense that $2|D| = S_+ \leq S_- = |D|\sqrt{2(\sqrt{2}+1)}$.

5 Polynomial equivalents of the shape function in higher degree

In degrees $m \geq 4$, we could not find analytical expressions of $K_{m,p}$ or $K_m^{\mathcal{E}}$, and do not expect them to exist. However, equivalent quantities with analytical expressions are available, under the same general form as in Theorem 4.1: the root of a polynomial in the coefficients of the polynomial $\pi \in \mathbb{H}_m$. This result improves on the analysis of [11], where a similar setting is studied.

In the following, we say that a function $\mathbf{R}$ is a polynomial on $\mathbb{H}_m$ if there exists a polynomial P of $m+1$ variables such that for all $(a_0, \cdots, a_m) \in \mathbb{R}^{m+1}$,

$$\mathbf{R}\left(\sum_{i=0}^{m} a_i x^i y^{m-i}\right) := P(a_0, \cdots, a_m),$$

and we define $\deg \mathbf{R} := \deg P$.

The object of this section is to prove the following theorem

Theorem 5.1 *For all degree $m \geq 2$, there exists a polynomial $\mathbf{K}_m$ on $I\!H_m$, and a constant $C_m > 0$ such that for all $\pi \in I\!H_m$, and all $1 \leq p \leq \infty$*

$$\frac{1}{C_m} \sqrt[r_m]{\mathbf{K}_m(\pi)} \leq K_{m,p}(\pi) \leq C_m \sqrt[r_m]{\mathbf{K}_m(\pi)},$$

where $r_m = \deg \mathbf{K}_m$.

Since for fixed m all functions $K_{m,p}$, $1 \leq p \leq \infty$, are equivalent on $\mathbb{H}_m$, there is no need to keep track of the exponent p in this section and we use below the notation $K_m = K_{m,\infty}$. In this section, please do not confuse the functions K_m and $\mathbf{K}_m$, as well as the polynomials Q_d and $\mathbf{Q}_d$ below, which notations are only distinguished by their case.

Theorem 5.1 is a generalization of Theorem 4.1, and the polynomial $\mathbf{K}_m$ involved should be seen as a generalization of the determinant on $\mathbb{H}_2$, and of the discriminant on $\mathbb{H}_3$. Let us immediately stress that the polynomial $\mathbf{K}_m$ is not unique. In particular, we shall propose two constructions that lead to different $\mathbf{K}_m$ with different degree r_m. Our first construction is simple and intuitive, but leads to a polynomial of degree r_m that grows quickly with m. Our second construction uses the tools of Invariant Theory to provide a polynomial of much smaller degree, which might be more useful in practice.

We first recall that there is a strong connection between the roots of a polynomial in $\mathbb{H}_2$ or $\mathbb{H}_3$ and its determinant or discriminant.

$$\det\left(\lambda\prod_{1\le i\le 2}(x-r_iy)\right) = \frac{-1}{4}\lambda^2(r_1-r_2)^2,$$
$$\operatorname{disc}\left(\lambda\prod_{1\le i\le 3}(x-r_iy)\right) = \lambda^4(r_1-r_2)^2(r_2-r_3)^2(r_3-r_1)^2.$$

We now fix an integer $m>3$. Observing that these expressions are a "cyclic" product of the squares of differences of roots, we define

$$\mathbb{S}(\lambda,r_1,\cdots,r_m) := \lambda^4(r_1-r_2)^2\cdots(r_{m-1}-r_m)^2(r_m-r_1)^2.$$

Since $m>3$, this quantity is not invariant anymore under reordering of the r_i. For any positive integer d, we introduce the symmetrized version of the d-powers of the cyclic product

$$Q_d(\lambda,r_1,\cdots,r_m) := \sum_{\sigma\in\Sigma_m}\mathbb{S}(\lambda,r_{\sigma_1},\cdots,r_{\sigma_m})^d,$$

where Σ_m is the set of all permutations of $\{1,\cdots,m\}$.

Proposition 5.1 *For all $d>0$ there exists a homogeneous polynomial $\mathbf{Q}_d$ of degree $4d$ on $\mathbb{H}_m$, with integer coefficients, and such that*

$$\textit{If } \pi=\lambda\prod_{i=1}^m(x-r_iy) \textit{ then } \mathbf{Q}_d(\pi)=Q_d(\lambda,r_1,\cdots,r_m).$$

In addition, $\mathbf{Q}_d$ obeys the invariance property

$$\mathbf{Q}_d(\pi\circ\phi) = (\det\phi)^{2md}\mathbf{Q}_d(\pi). \tag{39}$$

Proof: We denote by σ_i the elementary symmetric functions in the r_i, in such way that

$$\prod_{i=1}^m(x-r_iy) = x^m-\sigma_1x^{m-1}y+\sigma_2x^{m-2}y^2-\cdots+(-1)^m\sigma_my^m.$$

A well known theorem of algebra (see e.g. chapter IV.6 in [21]) asserts that any symmetrical polynomial in the r_i, can be reformulated as a polynomial in the σ_i. Hence for any d there exists a polynomial $\tilde{Q}_d$ such that

$$Q_d(1,r_1,\cdots,r_m)=\tilde{Q}_d(\sigma_1,\cdots,\sigma_m).$$

In addition it is known that the total degree of $\tilde{Q}_d$ is the partial degree of Q_d in the variable r_1, in our case $4d$, and that $\tilde{Q}_d$ has integer coefficients since Q_d has.

Given a polynomial $\pi\in H_m$ not divisible by y, we write it under the two equivalent forms

$$\pi=a_0x^m+a_1x^{m-1}y+\cdots+a_my^m=\lambda\prod_{i=1}^m(x-r_iy).$$

clearly $a_0=\lambda$ and $\sigma_i=(-1)^i\frac{a_i}{a_0}$. It follows that

$$Q_d(\lambda,r_1,\cdots,r_m)=\lambda^{4d}\tilde{Q}_d(\sigma_1,\cdots,\sigma_m)=a_0^{4d}\tilde{Q}_d\left(\frac{-a_1}{a_0},\cdots,\frac{(-1)^ma_m}{a_0}\right)$$

Since $\deg\tilde{Q}_d=4d$, the negative powers of a_0 due to the denominators are cleared by the factor a_0^{4d} and the right hand side is thus a polynomial in the coefficients $a_0,\cdots,a_m$ that we denote by $\mathbf{Q}_d(\pi)$.

We now prove the invariance of $\mathbf{Q}_d$ with respect to linear changes of coordinates, this proof is adapted from [19]. By continuity of $\mathbf{Q}_d$, it suffices to prove this invariance property for pairs (π, ϕ) such that ϕ is an invertible linear change of coordinates, and neither π or $\pi \circ \phi^{-1}$ is divisible by y.

Under this assumption, we observe that if $\pi = \lambda \prod_{i=1}^m (x - r_i y)$ and $\phi = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$, then $\pi \circ \phi^{-1} = \tilde{\lambda} \prod_{i=1}^m (x - \tilde{r}_i y)$ where

$$\tilde{\lambda} = \lambda (\det \phi)^{-m} \prod_{i=1}^m (\gamma + \delta r_i) \text{ and } \tilde{r}_i = \frac{\alpha r_i + \beta}{\gamma r_i + \delta}.$$

Observing that

$$\tilde{r}_i - \tilde{r}_j = \frac{\det \phi}{(\gamma r_i + \delta)(\gamma r_j + \delta)} (r_i - r_j),$$

it follows that

$$\mathbb{S}(\tilde{\lambda}, \tilde{r}_1, \cdots, \tilde{r}_m) = (\det \phi)^{-2m} \, \mathbb{S}(\lambda, r_1, \cdots, r_m).$$

The invariance property (39) follows readily. $\diamond$

We now define $r_m = 2\mathrm{lcm}\{\deg \mathbf{Q}_d \ : \ 1 \leq d \leq m!\}$ where $\mathrm{lcm}\{a_1, \cdots, a_k\}$ stands for the lowest common multiple of $\{a_1, \cdots, a_k\}$, and we consider the following polynomial on $\mathbb{H}_m$:

$$\mathbf{K}_m := \sum_{d=1}^{m!} \mathbf{Q}_d^{\frac{r_m}{\deg \mathbf{Q}_d}}.$$

Clearly $\mathbf{K}_m$ has degree r_m and obeys the invariance property $\mathbf{K}_m(\pi \circ \phi) = (\det \phi)^{\frac{r_m m}{2}} \mathbf{K}_m(\pi)$.

Lemma 5.2 *Let $\pi \in \mathbb{H}_m$. If $\mathbf{K}_m(\pi) = 0$ then $K_m(\pi) = 0$.*

Proof: We assume that $K_m(\pi) \neq 0$ and intend to prove that $\mathbf{K}_m(\pi) \neq 0$. Without loss of generality, we may assume that y does not divide π, since $K_m(\pi \circ U) = K_m(\pi)$ and $\mathbf{K}_m(\pi \circ U) = \mathbf{K}_m(\pi)$ for any rotation U. We thus write $\pi = \lambda \prod_{i=1}^m (x - r_i y)$, where $r_i \in \mathbb{C}$. Since $K_m(\pi) \neq 0$, we know from Proposition 2.1 that there is no group of $s_m := \lfloor \frac{m}{2} \rfloor + 1$ equal roots r_i.

We now define a permutation $\sigma^* \in \Sigma_m$ such that $r_{\sigma^*(i)} \neq r_{\sigma^*(i+1)}$ for $1 \leq i \leq m-1$ and $r_{\sigma^*(m)} \neq r_{\sigma^*(1)}$. In the case where $m = 2m'$ is even and m' of the r_i are equal, any permutation σ^* such that $r_{\sigma^*(1)} = r_{\sigma^*(3)} = \cdots = r_{\sigma^*(2m'-1)}$ satisfies this condition. In all other cases let us assume that the r_i are sorted by equality : if $i < j < k$ and $r_i = r_k$ then $r_i = r_j = r_k$. If $m = 2m'$ is even, we set $\sigma^*(2i-1) = i$ and $\sigma^*(2i) = m' + i$, $1 \leq i \leq m'$. If $m = 2m'+1$ is odd we set $\sigma^*(2i) = i$, $1 \leq i \leq m'$ and $\sigma^*(2i-1) = m' + i$, $1 \leq i \leq m'+1$. For example, $\sigma^* = (4\ 1\ 5\ 2\ 6\ 3\ 7)$ when $m = 7$ and $\sigma^* = (1\ 5\ 2\ 6\ 3\ 7\ 4\ 8)$ when $m = 8$. With such a construction, we find that $|\sigma^*(i) - \sigma^*(i+1)| \geq m'$ if m is odd and $|\sigma^*(i) - \sigma^*(i+1)| \geq m' - 1$ if m is even, for all $1 \leq i \leq m$, where we have set $\sigma^*(m+1) := \sigma^*(1)$. Hence σ satisfies the required condition, and therefore $\mathbb{S}(\lambda, r_{\sigma^*(1)}, \cdots, r_{\sigma^*(m)}) \neq 0$.

It is well known that if k complex numbers $\alpha_1, \cdots, \alpha_k \in \mathbb{C}$ are such that $\alpha_1^d + \cdots + \alpha_k^d = 0$, for all $1 \leq d \leq k$, then $\alpha_1 = \cdots = \alpha_k = 0$. Applying this property to the $m!$ complex numbers $\mathbb{S}(\lambda, r_{\sigma(1)}, \cdots, r_{\sigma(m)})$, $\sigma \in \Sigma_m$, and noticing that the term corresponding to σ^* is non zero, we see that there exists $1 \leq d \leq m!$ such that $\mathbf{Q}_d(\pi) = Q_d(\lambda, r_1, \cdots, r_m) \neq 0$. Since $\mathbf{Q}_d$ has real coefficients, the numbers $\mathbf{Q}_d(\pi)$ are real. Since the exponent $r_m / \deg \mathbf{Q}_d$ is even it follows that $\mathbf{K}_m(\pi) > 0$, which concludes the proof of this lemma. $\diamond$

The following proposition, when applied to the function $K_{\mathrm{eq}} = \sqrt[r_m]{\mathbf{K}_m}$ concludes the proof of Theorem 5.1.

Proposition 5.3 *Let $m \geq 2$, and let $K_{\mathrm{eq}} : \mathbb{H}_m \to \mathbb{R}_+$ be a continuous function obeying the following properties*

1. Invariance property : $K_{\mathrm{eq}}(\pi \circ \phi) = |\det \phi|^{\frac{m}{2}} K_{\mathrm{eq}}(\pi)$.

2. Vanishing property *: for all $\pi \in \mathbb{H}_m$, if $K_{\mathrm{eq}}(\pi) = 0$ then $K_m(\pi) = 0$.*

Then there exists a constant $C > 0$ such that $\frac{1}{C} K_{\mathrm{eq}} \le K_m \le C K_{\mathrm{eq}}$ on $\mathbb{H}_m$.

Proof: We first remark that K_{eq} is homogeneous in a similar way as K_m: if $\lambda \ge 0$, then applying the invariance property to $\phi = \lambda^{\frac{1}{m}} \mathrm{Id}$ yields $K_{\mathrm{eq}}(\pi \circ (\lambda^{\frac{1}{m}} \mathrm{Id})) = K_{\mathrm{eq}}(\lambda \pi)$ and $|\det \phi|^{\frac{m}{2}} = \lambda$. Hence $K_{\mathrm{eq}}(\lambda \pi) = \lambda K_{\mathrm{eq}}(\pi)$.

Our next remark is that a converse of the vanishing property holds: if $K_m(\pi) = 0$, then there exists a sequence ϕ_n of linear changes of coordinates, $\det \phi_n = 1$, such that $\pi \circ \phi_n \to 0$ as $n \to \infty$. Hence $K_{\mathrm{eq}}(\pi) = K_{\mathrm{eq}}(\pi \circ \phi_n) \to K_{\mathrm{eq}}(0)$. Furthermore, $K_{\mathrm{eq}}(0) = 0$ by homogeneity. Hence $K_{\mathrm{eq}}(\pi) = 0$.

We define the set $NF_m := \{\pi \in \mathbb{H}_m \ : \ K_m(\pi) = 0\}$. We also define a set $A_m \subset \mathbb{H}_m$ by a property "opposite" to the property defining NF_m. A polynomial $\pi \in \mathbb{H}_m$ belongs to A_m if and only if

$$\|\pi\| \le \|\pi \circ \phi\| \text{ for all } \phi \text{ such that } \det \phi = 1.$$

The sets NF_m and A_m are closed by construction, and clearly $NF_m \cap A_m = \{0\}$. We now define

$$\underline{K_m}(\pi) = \lim_{r \to 0} \inf_{\|\pi' - \pi\| \le r} K_m(\pi')$$

the lower semi-continuous envelope of K_m. If $\underline{K_m}(\pi) = 0$ then there exists a converging sequence $\pi_n \to \pi$ such that $K_m(\pi_n) \to 0$. According to Proposition 2.3, it follows that $K_m(\pi) = 0$ and hence $\pi \in NF_m$. Therefore the lower semi continuous function $\underline{K_m}$ and the continuous function K_{eq} are bounded below by a positive constant on the compact set $\{\pi \in A_m, \|\pi\| = 1\}$. Since in addition K_{eq} is continuous and K_m is upper semi-continuous, we find that the constant

$$C = \sup_{\pi \in A_m, \|\pi\| = 1} \max \left\{ \frac{K_{\mathrm{eq}}(\pi)}{\underline{K_m}(\pi)}, \frac{K_m(\pi)}{K_{\mathrm{eq}}(\pi)} \right\},$$

is finite. By homogeneity of K_m and K_{eq}, we infer that on A_m

$$\frac{1}{C} K_{\mathrm{eq}} \le \underline{K_m} \le K_m \le C K_{\mathrm{eq}}. \tag{40}$$

Now, for any $\pi \in \mathbb{H}_m$, we consider $\hat{\pi}$ of minimal norm in the closure of the set $\{\pi \circ \phi \ : \ \det \phi = 1\}$. By construction, we have $\hat{\pi} \in A_m$, and there exists a sequence ϕ_n, $\det \phi_n = 1$ such that $\pi \circ \phi_n \to \hat{\pi}$ as $n \to \infty$. If $\hat{\pi} = 0$, then $K_m(\pi) = K_{\mathrm{eq}}(\pi) = 0$. Otherwise, we observe that

$$\underline{K_m}(\hat{\pi}) \le K_m(\pi) \le K_m(\hat{\pi}) \text{ and } K_{\mathrm{eq}}(\hat{\pi}) = K_{\mathrm{eq}}(\pi).$$

Where we used the fact that $\underline{K_m}$, K_m and K_{eq} are respectively lower semi continuous, upper semi continuous, and continuous on $\mathbb{H}_m$. Combining this with inequality (40) concludes the proof. $\diamond$

A natural question is to find the polynomial of smallest degree satisfying Theorem 5.1. This leads us to the theory of *invariant polynomials* introduced by Hilbert [19] (we also refer to [16] for a survey on this subject). A polynomial R on $\mathbb{H}_m$ is said to be invariant if $\mu = \frac{m \deg R}{2}$ is a positive integer and for all $\pi \in \mathbb{H}_m$ and linear change of coordinates ϕ, one has

$$R(\pi \circ \phi) = (\det \phi)^\mu R(\pi). \tag{41}$$

We have seen for instance that $\mathbf{K}_m$ and $\mathbf{Q}_d$ are "invariant polynomials" on $\mathbb{H}_m$.

Nearly all the literature on invariant polynomials is concerned with the case of complex coefficients, both for the polynomials and the changes of variables. It is known in particular [16] that for all $m \ge 3$, there exists $m-2$ invariant polynomials $R_1, \cdots R_{m-2}$ on $\mathbb{H}_m$, such that for any π (complex coefficients are allowed) and any other invariant polynomial R on $\mathbb{H}_m$,

$$\text{If } R_1(\pi) = \cdots = R_{m-2}(\pi) = 0, \text{ then } R(\pi) = 0. \tag{42}$$

A list of such polynomials with minimal degree is known explicitly at least when $m \leq 8$. Defining $r = 2\mathrm{lcm}(\deg R_i)$ and $K_{\rm eq} := \sqrt[r]{\sum_{i=1}^{m-2} \mathbf{R}_i^{\frac{r}{\deg R_i}}}$, we see that $K_{\rm eq}(\pi) = 0$ implies $\mathbf{K}_m(\pi) = 0$ and hence $K_m(\pi) = 0$. According to proposition 5.3, we have constructed a new, possibly simpler, equivalent of K_m.

For example when $m = 2$ the list (R_i) is reduced to the polynomial det, and for $m = 3$ to the polynomial disc. For $m = 4$, given $\pi = ax^4 + 4bx^3y + 6cx^2y^2 + 4dxy^3 + ey^4$, the list consists of the two polynomials

$$I = ae - 4bd + 3c^2, \quad J = \begin{vmatrix} a & b & c \\ b & c & d \\ c & d & e \end{vmatrix},$$

therefore $K_4(\pi)$ is equivalent to the quantity $\sqrt[6]{|I(\pi)|^3 + J(\pi)^2}$. As m increases these polynomials unfortunately become more and more complicated, and their number $m-2$ obviously increases. According to [16], for $m = 5$ the list consists of three polynomials of degrees $4, 8, 12$, while for $m = 6$ it consists of 4 polynomials of degrees $2, 4, 6, 10$.

6 Extension to higher dimension

The function $K_{m,p}$ can be generalized to higher dimension $d > 2$ in the following way. We denote by $\mathbb{H}_{m,d}$ the set of homogeneous polynomials of degree m in d variables. For all d-dimensional simplex T, we define the interpolation operator $I_{m,T}$ acting from $C^0(T)$ onto the space $\mathbb{P}_{m-1,d}$ of polynomials of total degree $m-1$ in d variables. This operator is defined by the conditions $I_{mT}v(\gamma) = v(\gamma)$ for all point $\gamma \in T$ with barycentric coordinates in the set $\{0, \frac{1}{m-1}, \frac{2}{m-1}, \cdots, 1\}$. Following Section §1.2, and generalizing Definition (2), we define the local interpolation error on a simplex, the global interpolation error on a mesh, as well as the shape function.

For all $\pi \in \mathbb{H}_{m,d}$,

$$K_{m,p,d}(\pi) := \inf_{|T|=1} \|\pi - I_{m,T}\pi\|_p.$$

where the infimum is taken on all d-dimensional simplexes T of volume 1. The variant $K_m^{\mathcal{E}}$ introduced in (14) also generalizes in higher dimension, and was introduced by Weiming Cao in [9]. Denoting by $\mathcal{E}_d$ the set of d-dimensional ellipsoids, we define

$$K_{m,d}^{\mathcal{E}}(\pi) = \left(\sup_{E \in \mathcal{E}_d, E \subset \Lambda_\pi} |E|\right)^{-\frac{m}{d}},$$

with $\Lambda_\pi = \{z \in \mathbb{R}^d \ : \ |\pi(z)| \leq 1\}$. Similarly to Proposition 2.4, it is not hard to show that the functions $K_{m,p,d}(\pi)$ and $K_{m,d}^{\mathcal{E}}(\pi)$ are equivalent: there exists constants $0 < c \leq C$ depending only on m, d, such that

$$cK_{m,d}^{\mathcal{E}} \leq K_{m,p,d} \leq CK_{m,d}^{\mathcal{E}}.$$

Let $(\mathcal{T}_n)_{n\geq 0}$ be a sequence of simplicial meshes (triangles if $d = 2$, tetrahedrons if $d = 3$, ...) of a d-dimensional, polygonal open set Ω. Generalizing (5), we say that $(\mathcal{T}_n)_{n\geq 0}$ is admissible if there exists a constant C_A verifying

$$\sup_{T \in \mathcal{T}_n} \mathrm{diam}(T) \leq C_A N^{-1/d}.$$

The lower estimate in Theorem 1.2 can be generalized, with straightforward adaptations in the proof. If $f \in C^m(\Omega)$ and $(\mathcal{T}_N)_{N \geq N_0}$ is an admissible sequence of triangulations, then

$$\liminf_{N \to \infty} N^{\frac{m}{d}} e_{m,\mathcal{T}_N}(f)_p \geq \left\| K_{m,d,p}\left(\frac{d^m f}{m!}\right)\right\|_{L^q(\Omega)}.$$

Where $\frac{1}{q} := \frac{m}{d} + \frac{1}{p}$.

The upper estimate in Theorem 1.2 however does not generalize. The reason is that we used in its proof a tiling of the plane consisting of translates of a single triangle and of its symmetric with respect to the origin. This construction is not possible anymore in higher dimension, for example it is well known that one cannot tile the space $\mathbb{R}^3$, with equilateral tetrahedra.

The generalization of the second part of Theorem (1.2) is therefore the following. For all m and d, there exists a constant $C = C(m,d) > 0$, such that for any polygonal open set $\Omega \subset \mathbb{R}^d$ and $f \in C^m(\Omega)$ the following holds: for all $\varepsilon > 0$, there exists an admissible sequence $\mathcal{T}_n$ of triangulations of Ω such that

$$\limsup_{N\to\infty} N^{\frac{m}{d}} e_{m,\mathcal{T}_N}(f)_p \le C \left\| K_{m,d,p}\left(\frac{d^m f}{m!}\right)\right\|_{L^q(\Omega)} + \varepsilon.$$

The "tightness" Theorem 1.2 is partially lost due to the constant C. This upper bound is not new, and can be found in [9]. In the proof of the bidimensional theorem we define by (23) a tiling $\mathcal{P}_R$ of the plane made of a triangle T_R, and some of its translates and of their symmetry with respect to the origin. In dimension d, the tiling $\mathcal{P}_R$ cannot be constructed by the same procedure. The idea of the proof is to first consider a fixed tiling $\mathcal{P}_0$ of the space, constituted of simplices bounded diameter, and of volume bounded below by a positive constant, as well as a reference equilateral simplex $T_{\rm eq}$ of volume 1. We then set $\mathcal{P}_R = \phi(\mathcal{P}_0)$, where ϕ is a linear change of coordinates such that $T_R = \phi(T_{\rm eq})$. This procedure can be applied in any dimension, and yields all subsequent estimates "up to a multiplicative constant", which concludes the proof.

Since this upper bound is not tight anymore, and since the functions $K_{m,p,d}$ are all equivalent to $K^{\mathcal{E}}_{m,d}$ as p varies (with equivalence constants independent of p), there is no real need to keep track of the exponent p. We therefore denote by $K_{m,d}$ the function $K_{m,\infty,d}$.

For practical as well as theoretical purposes, it is desirable to have an efficient way to compute the shape function $K_{m,d}$, and an efficient algorithm to produce adapted triangulations. The case $m = 2$, which corresponds to piecewise linear elements, has been extensively studied see for instance [4, 12]. In that case there exists constants $0 < c < C$, depending only on d, such that for all $\pi \in \mathbb{H}_{2,d}$,

$$c\sqrt[d]{|\det\pi|} \le K_{2,d}(\pi) \le C\sqrt[d]{|\det\pi|}.$$

where $\det\pi$ denotes the determinant of the symmetric matrix associated to π. Furthermore, similarly to Proposition 4.2, the optimal metric for mesh refinement is given by the absolute value of the matrix of second derivatives, see [4, 12], which is constructed in a similar way as in dimension $d = 2$: with U and $D = \mathrm{diag}(\lambda_1, \cdots, \lambda_d)$ the orthogonal and diagonal matrices such that $[\pi] = U^{\mathrm{T}}DU$ and with $|D| := \mathrm{diag}(|\lambda_1|, \cdots, |\lambda_d|)$, we set $h_\pi = U^{\mathrm{T}}|D|U$. It can be shown that the matrix h_π defines an ellipsoid of maximal volume included into the set Λ_π. The case $m = 2$ can therefore be regarded as solved.

For values (m,d) both larger than 2, the question of computing the shape function as well as the optimal metric is much more difficult, but we have partial answers, in particular for quadratic elements in dimension 3. Following §5, we need fundamental results from the theory of invariant polynomials, developed in particular by Hilbert [19]. In order to apply these results to our particular setting, we need to introduce a compatibility condition between the degree m and the dimension d.

Definition 6.1 *We call the pair of numbers $m \ge 2$ and $d \ge 2$ "compatible" if and only if the following holds. For all $\pi \in \mathbb{H}_{m,d}$ such that there exists a sequence $(\phi_n)_{n\ge 0}$ of $d\times d$ matrices with* complex *coefficients, verifying $\det\phi_n = 1$ and $\lim_{n\to\infty}\pi\circ\phi_n = 0$, there also exists a sequence ψ_n of $d\times d$ matrices with* real *coefficients, verifying $\det\psi_n = 1$ and $\lim_{n\to\infty}\pi\circ\psi_n = 0$.*

Following Hilbert [19], we say that a polynomial Q of degree r defined on $\mathbb{H}_{m,d}$ is *invariant* if $\mu = \frac{mr}{d}$ is a positive integer and if for all $\pi \in \mathbb{H}_{m,d}$ and all linear changes of coordinates ϕ,

$$Q(\pi\circ\phi) = (\det\phi)^\mu Q(\pi). \tag{43}$$

This is a generalization of (41). We denote by $\mathbb{I}_{m,d}$ the set of invariant polynomials on $\mathbb{H}_{m,d}$. It is easy to see that if $\pi \in \mathbb{H}_{m,d}$ is such that $K_{m,d}(\pi) = 0$, then $Q(\pi) = 0$ for all $Q \in \mathbb{I}_{m,d}$. Indeed, as seen in the proof of Proposition 2.1, if $K_{m,d}(\pi) = 0$ then there exists a sequence ϕ_n such that $\det\phi_n = 1$ and $\pi\circ\phi_n \to 0$. Therefore (43) implies that $Q(\pi) = 0$. The following lemma shows that the compatibility condition for the pair (m,d) is equivalent to a converse of this property.

Lemma 6.1 *The pair (m,d) is compatible if and only if for all $\pi \in \mathbb{H}_{m,d}$*

$$K_{m,d}(\pi) = 0 \text{ if and only if } Q(\pi) = 0 \text{ for all } Q \in \mathbb{I}_{m,d}.$$

Proof: We first assume that the pair (m,d) is not compatible. Then there exists a polynomial $\pi_0 \in \mathbb{H}_{m,d}$ such that there exists a sequence ϕ_n, $\det \phi_n = 1$ of matrices with *complex* coefficients such that $\pi \circ \phi_n \to 0$, but there exists no such sequence with *real* coefficients. This last property indicates that $K_{m,d}(\pi) > 0$. On the contrary let $Q \in \mathbb{I}_{m,d}$ be an invariant polynomial, and set $\mu = \frac{m \deg Q}{d}$. The identity

$$Q(\pi_0 \circ \phi) = (\det \phi)^\mu Q(\pi_0)$$

is valid for all ϕ with real coefficients, and is a *polynomial identity* in the coefficients of ϕ. Therefore it remains valid if ϕ has complex coefficients. If follows that $Q(\pi_0) = Q(\pi_0 \circ \phi_n)$ for all n, and therefore $Q(\pi_0) = 0$, which concludes the proof in the case where the pair (m,d) is not compatible.

We now consider a compatible pair (m,d). Following Hilbert [19], we say that a polynomial $\pi \in H_{m,d}$ is a *null form* if and only if there exists a sequence of matrices ϕ_n with *complex* coefficients such that $\det \phi_n = 1$ and $\pi \circ \phi_n \to 0$. We denote by $NF_{m,d}$ the set of such polynomials. Since the pair (m,d) is compatible, note that $\pi \in NF_{m,d}$ if and only if there exists a sequence ϕ_n of matrices with *real* coefficients such that $\det \phi_n = 1$ and $\pi \circ \phi_n \to 0$. Hence, we find that

$$NF_{m,d} = \{\pi \in \mathbb{H}_{m,d} \ : \ K_{m,d}(\pi) = 0\}.$$

Denoting by $\mathbb{I}_{m,d}^{\mathbb{C}}$ the set of invariant polynomials on $\mathbb{H}_{m,d}$ with *complex* coefficients, a difficult theorem of [19] states that

$$NF_{m,d} = \{\pi \in \mathbb{H}_{m,d} \ : \ Q(\pi) = 0 \text{ for all } Q \in \mathbb{I}_{m,d}^{\mathbb{C}}\}$$

It is not difficult to check that if $Q = Q_1 + iQ_2$ where Q_1 and Q_2 have real coefficients then (43) holds for Q if and only if it holds for both Q_1 and Q_2, i.e. Q_1 and Q_2 are also invariant polynomials. Hence denoting by $\mathbb{I}_{m,d}$ the set of invariant polynomials on $\mathbb{H}_{m,d}$ with real coefficients, we have obtained that

$$NF_{m,d} = \{\pi \in \mathbb{H}_{m,d} \ : \ Q(\pi) = 0 \text{ for all } Q \in \mathbb{I}_{m,d}\}$$

which concludes the proof. $\diamond$

Theorem 6.1 *If the pair (m,d) is compatible, then there exists a polynomial $\mathbf{K}$ on $\mathbb{H}_{m,d}$ (we set $r = \deg \mathbf{K}$) and a constant $C > 0$ such that for all $\pi \in \mathbb{H}_{m,d}$*

$$\frac{1}{C} \sqrt[r]{\mathbf{K}(\pi)} \le K_{m,d}(\pi) \le C \sqrt[r]{\mathbf{K}(\pi)}. \tag{44}$$

If the pair (m,d) is not compatible, then there does not exist such a polynomial $\mathbf{K}$.

Proof: The proof of the non-existence property when the pair (m,d) is not compatible is reported in the appendix. Assume that the pair (m,d) is compatible. We follow a reasoning very similar to §5 to prove the equivalence (44).

We use the notations of Lemma 6.1 and consider the set

$$NF_{m,d} = \{\pi \in \mathbb{H}_{m,d} \ : \ K_{m,d}(\pi) = 0\} = \{\pi \in \mathbb{H}_{m,d} \ : \ Q(\pi) = 0, \ Q \in \mathbb{I}_{m,d}\}.$$

The ring of polynomials on a field is known to be Noetherian. This implies that there exists a finite family $Q_1, \cdots, Q_s \in \mathbb{I}_{m,d}$ of invariant polynomials on $\mathbb{H}_{m,d}$ such that any invariant polynomial is of the form $\sum P_i Q_i$ where P_i are polynomials on $\mathbb{H}_{m,d}$. We therefore obtain

$$NF_{m,d} = \{\pi \in \mathbb{H}_{m,d} \ : \ Q_1(\pi) = \cdots = Q_s(\pi) = 0\}.$$

which is a generalization of (42), however with no clear bound on s.

We now fix such a set of polynomials, set $r := 2\mathrm{lcm}_{1 \le i \le s} \deg Q_i$, and define

$$\mathbf{K} = \sum_{i=1}^{s} \mathbf{Q}_i^{\frac{r}{\deg \mathbf{Q}_i}} \ \text{ and } \ K_{\mathrm{eq}} := \sqrt[r]{\mathbf{K}}.$$

Clearly $\mathbf{K}$ is an invariant polynomial on $\mathbb{H}_{m,d}$, and $NF_{m,d} = \{\pi \in \mathbb{H}_{m,d} \ : \ \mathbf{K}(\pi) = 0\}$. Hence the function $K_{\rm eq}$ is *continuous* on $\mathbb{H}_{m,d}$, obeys the invariance property $K_{\rm eq}(\pi \circ \phi) = |\det \phi| K_{\rm eq}(\pi)$, and for all $\pi \in \mathbb{H}_m$, $K_{\rm eq}(\pi) = 0$ implies $\mathbf{K}(\pi) = 0$ and therefore $K_{m,d}(\pi) = 0$. We recognize here the hypotheses of Proposition 5.3, except that the dimension d has changed. Inspection of the proof of Proposition 5.3 shows that we use only once the fact that $d = 2$, when we refer to Proposition 2.3 and state that if $(\pi_n) \in \mathbb{H}_m$, $\pi_n \to \pi$ and $K_m(\pi_n) \to 0$, then $K_m(\pi) = 0$. This property also applies to $K_{m,d}$, when the pair (m, d) is compatible. Assume that $(\pi_n) \in \mathbb{H}_{m,d}$, $\pi_n \to \pi$ and that $K_{m,d}(\pi_n) \to 0$. Then there exists a sequence of linear changes of coordinates ϕ_n, $\det \phi_n = 1$, such that $\pi_n \circ \phi_n \to 0$. Therefore

$$\mathbf{K}(\pi) = \lim_{n\to\infty} \mathbf{K}(\pi_n) = \lim_{n\to\infty} \mathbf{K}(\pi_n \circ \phi_n) = 0$$

It follows that $\pi \in NF_{m,d}$, and therefore $K_{m,d}(\pi) = 0$. Since the rest of the proof of Proposition 5.3 never uses that $d = 2$, this concludes the proof of Equivalence (44). $\diamond$

Hence there exists a "simple" equivalent of $K_{m,d}$ for all compatible pairs (m, d), while equivalents of $K_{m,d}$ for incompatible pairs need to be more sophisticated, or at least different from the root of a polynomial. This theorem leaves open several questions. The first one is to identify the list of compatible pairs (m, d). It is easily shown that the pairs $(m, 2)$, $m \geq 2$, and $(2, d)$, $d \geq 2$ are compatible, but this does not provide any new results since we already derived equivalents of the shape function in these cases. More interestingly, we show in the next corollary that the pair $(3, 3)$ is compatible, which corresponds to approximation by quadratic elements in dimension 3. There exists two generators S and T of $I_{3,3}$, which expressions are given in [23] and which have respectively degree 4 and 6.

Corollary 6.1 *$\sqrt[6]{|S|^3 + T^2}$ is equivalent to $K_{3,3}$ on $\mathbb{H}_{3,3}$.*

Proof: The invariants S and T obey the invariance properties $S(\pi \circ \phi) = (\det \phi)^4 S(\pi)$ and $T(\pi \circ \phi) = (\det \phi)^6 T(\pi)$. We intend to show that if $\pi \in \mathbb{H}_{3,3}$ and $S(\pi) = T(\pi) = 0$ then $K_{3,3}(\pi) = 0$. Let us first admit this property and see how to conclude the proof of this corollary. According to Lemma 6.1 the pair $(3, 3)$ is compatible. The function $K_{\rm eq} := \sqrt[6]{|S|^3 + T^2}$ is continuous on $\mathbb{H}_{3,3}$, obeys the invariance property $K_{\rm eq}(\pi \circ \phi) = |\det \phi| K_{\rm eq}(\pi)$ and is such that $K_{\rm eq}(\pi) = 0$ implies $K_{3,3}(\pi) = 0$. We have seen in the proof of Theorem 6.1 that these properties imply the desired equivalence of $K_{\rm eq}$ and $K_{3,3}$.

We now show that $S(\pi) = T(\pi) = 0$ implies $K_{3,3}(\pi) = 0$. A polynomial $\pi \in \mathbb{H}_{3,3}$ can be of two types. Either it is *reducible*, meaning that there exists $\pi_1 \in \mathbb{H}_{1,3}$ (linear) and $\pi_2 \in \mathbb{H}_{2,3}$ (quadratic) such that $\pi = \pi_1\pi_2$, or it is *irreducible*. In the latter case according to [18], there exists a linear change of coordinates ϕ and two reals a, b such that

$$\pi \circ \phi = y^2 z - (x^3 + 3axz^2 + bz^3).$$

A direct computation from the expressions given in [23] shows that $S(\pi \circ \phi) = a$ and $T(\pi \circ \phi) = -4b$. If $S(\pi) = T(\pi) = 0$ then $S(\pi \circ \phi) = T(\pi \circ \phi) = 0$ and $\pi \circ \phi = y^2 z - x^3$. Therefore for all $\lambda \neq 0$, $\pi \circ \phi(\lambda x, \lambda^2 y, \lambda^{-3} z) = \lambda y^2 z - \lambda^3 x^3$, which tends to 0 as $\lambda \to 0$. We easily construct from this point a sequence ϕ_n, $\det \phi_n = 1$, such that $\pi \circ \phi_n \to 0$. Therefore $K_{3,3}(\pi) = 0$.

If π is reducible, then $\pi = \pi_1\pi_2$ where π_1 is linear and π_2 is quadratic. Choosing a linear change of coordinates ϕ such that $\pi_1 \circ \phi = z$ we obtain

$$\pi \circ \phi = 3z(ax^2 + 2bxy + cy^2) + z^2(ux + vy + wz),$$

for some constants a, b, c, u, v, w. Again, a direct computation from the expressions given in [23] shows that $S(\pi \circ \phi) = -(ac - b^2)^2$ (and $T(\pi \circ \phi) = 8(ac - b^2)^3$). Therefore if $S(\pi) = T(\pi) = 0$ then the quadratic function $ax^2 + 2bxy + cy^2$ of the pair of variables (x, y) is degenerate. Hence there exists a linear change of coordinates ψ, altering only the variables x, y, and reals μ, u', v' such that

$$\pi \circ \phi \circ \psi = \mu z x^2 + z^2(u'x + v'y + wz).$$

It follows that $\pi \circ \phi \circ \psi(x, \lambda^{-1} y, \lambda z)$ tends to 0 as $\lambda \to 0$. Again, this implies that $K_{3,3}(\pi) = 0$, and concludes the proof of this proposition. $\diamond$

We could not find any example of incompatible pair (m, d), which leads us to formulate the conjecture that all pairs (m, d) are compatible (hence providing "simple" equivalents of $K_{m,d}$ in full generality). Another even more difficult problem is to derive a polynomial $\mathbf{K}$ of minimal degree for all couples (m, d) which are compatible and of interest.

Last but not least, efficient algorithms are needed to compute metrics, from which effective triangulations are built that yield the optimal estimates. A possibility is to follow the approach proposed in [9], i.e. solve numerically the optimization problem

$$\inf\{\det H \ : \ H \in S_d^+ \text{ and } \forall z \in \mathbb{R}^d, \langle Hz, z\rangle \geq |\pi(z)|^{2/m}\},$$

which amounts to minimizing a degree d polynomial under an infinite set of linear constraints. When $d > 2$, this minimization problem is not quadratic which makes it rather delicate. Furthermore, numerical instabilities similar to those described in Remark 4.4 can be expected to appear.

7 Conclusion and Perspectives

In this paper, we have introduced asymptotic estimates for the finite element interpolation error measured in L^p when the mesh is optimally adapted to the interpolated function. These estimates are asymptotically sharp for functions of two variables, see Theorem 1.2, and precise up to a fixed multiplicative constant in higher dimension, as described in §6. They involve a shape function $K_{m,p}$ (or $K_{m,d,p}$ if $d > 2$) which generalizes the determinant which appears in estimates for piecewise linear interpolation [12, 4, 13]. This function can be explicitly computed in several cases, as shows Theorem 4.1, and has equivalents of a simple form in a number of other cases, see Theorems 5.1 and 6.1.

All our results are stated and proved for sufficiently smooth functions. One of our future objectives is to extend these results to larger classes of functions, and in particular to functions exhibiting discontinuities along curves. This means that we need to give a proper meaning to the nonlinear quantity $K_{m,p}\left(\frac{d^m f}{m!}\right)$ for non-smooth functions.

This paper also features a constructive algorithm (similar to [4]), that produces triangulations obeying our sharp estimates, and is described in §3.2. However, this algorithm becomes asymptotically effective only for a highly refined triangulation. A more practical way to produce quasi-optimal triangulations is to adapt them to a metric, see [6, 20, 7]. This approach is discussed in §4.2. This raises the question of generating the appropriate metric from the (approximate) knowledge of the derivatives of the function to be interpolated. We addressed this question in the particular case of piecewise quadratic approximation in two dimensions in Theorems 4.2 and 4.3.

We plan to integrate this result in the PDE solver FreeFem++ in a near future. Note that a Mathematica source code is already available on the web [25]. We also would like to derive appropriate metrics for other settings of degree m and dimension d, although, as we pointed it in Proposition 4.4, this might be a rather delicate matter.

We finally remark that in many applications, one seeks for error estimates in the Sobolev norms $W^{1,p}$ (or $W^{m,p}$) rather than in the L^p norms. Finding the optimal triangulation for such norms requires a new error analysis. For instance, in the survey [24] on piecewise linear approximation, it is observed that the metric $h_\pi = |d^2 f|$ (evoked in Equation (37)) should be replaced with $h_\pi = (d^2 f)^2$ for best adaptation in H^1 norm. In other words, the principal axes of the positive definite matrix h_π remain the same, but its conditioning is squared.

APPENDIX

A Proof of Proposition 4.3

We consider a fixed polynomial $\pi \in \mathbb{H}_3$, a parameter $\alpha > 0$, and look for an ellipse $E_{\pi,\alpha}$ of maximal volume included in the set $\alpha^{-1/2} D \cap \Lambda_\pi$. Since this set is compact, a standard argument shows that there exists at least one such ellipse.

If $\alpha \geq \mu_\pi$, then $\alpha^{-1/2} D \subset \Lambda_\pi$ and therefore $\alpha^{-1/2} D \cap \Lambda_\pi = \alpha^{-1/2} D$. It follows that $E_{\pi,\alpha} = \alpha^{-1/2} D$, which proves part 1.

In the following we denote by E'_α the ellipse defined by the matrix (38). Note that any ellipse containing D_π and included in Λ_π must be tangent to $\partial\Lambda_\pi$ at the point z_π, and hence of the form E'_δ for some $\delta > 0$. Clearly $E'_\delta \subset E'_\mu$ if and only if $\delta \geq \mu$. Therefore $E'_\alpha \subset \Lambda_\pi$ if and only if $\alpha \geq \alpha_\pi$. Let E be an arbitrary ellipse, let D_1 the largest disc contained in E, and D_2 the smallest disc containing E. Then it is not hard to check that $|E| = \sqrt{|D_1||D_2|}$. For any α verifying $\alpha_\pi \leq \alpha \leq \mu_\pi$, the ellipse E'_α is such that $D_1 = D_\pi$, which is the largest centered disc contained in Λ_π, and $D_2 = \alpha^{-1/2}D$, which corresponds to the bound $2\alpha^{-1/2}$ on the diameter of $E_{\pi,\alpha}$. It follows that E'_α is an ellipse of maximal volume included in $\alpha^{-1/2}D \cap \Lambda_\pi$, and this concludes the proof of part 2.

Part 4 is trivial, hence we concentrate on part 3 and assume that $\beta_\pi \leq \alpha \leq \alpha_\pi$.

An elementary observation is that $E_{\pi,\alpha}$ must be "blocked with respect to rotations". Indeed assume for contradiction that $R_\theta(E_{\pi,\alpha}) \subset \Lambda_\pi$ for $\theta \in [0, \varepsilon]$ or $[-\varepsilon, 0]$, where we denote by R_θ the rotation of angle θ. Observing that the set $\cup_{\theta\in[0,\varepsilon]} R_\theta(E_{\pi,\alpha})$ contains an ellipse of larger area than $E_{\pi,\alpha}$ and of the same diameter, we obtain a contradiction.

In the following, we say that an ellipse E is quadri-tangent to Λ_π, when there are at least four points of tangency between ∂E and $\partial\Lambda_\pi$ (a tangency point being counted twice if the radii of curvature of ∂E and $\partial\Lambda_\pi$ coincide at this point).

The fact that $E_{\pi,\alpha}$ is "blocked with respect to rotations" implies that it is either quadri-tangent to Λ_π or tangent to $\partial\Lambda_\pi$ at the extremities of its small axis. In the latter case the extremities of the small axis must clearly be the points z_π and $-z_\pi$, the closest points of $\partial\Lambda_\pi$ to the origin. It follows that $E_{\pi,\alpha}$ belongs to the family E'_δ, $\delta \geq \alpha_\pi$ described above, and therefore is equal to E'_{α_π} since $\alpha \leq \alpha_\pi$. But E'_{α_π} is quadri-tangent to Λ_π, since otherwise we would have $E'_{\alpha_\pi-\varepsilon} \subset \Lambda_\pi$ for some $\varepsilon > 0$.

We have now established that $E_{\pi,\alpha}$ is quadri-tangent to Λ_π when $\beta_\pi \leq \alpha \leq \alpha_\pi$. This property is invariant by any linear change of coordinate: if an ellipse E is quadri-tangent to $\Lambda_{\pi\circ\phi}$, then $\phi(E)$ is quadri-tangent to Λ_π. Furthermore if E is defined by a symmetric positive definite matrix H, then $\phi(E)$ is defined by $(\phi^{-1})^{\mathrm{T}} H \phi^{-1}$. This remark leads us to identify the family of ellipses quadri-tangent to $\partial\Lambda_\pi$ when π is among the four reference polynomials $x(x^2 - 3y^2)$, $x(x^2 + 3y^2)$, x^2y and x^3. In the case of x^3 there is no quadri-tangent ellipse and we have $\alpha_\pi = 0$, therefore part 3 of the theorem is irrelevant. In the three other cases, which respectively correspond to part 3 (i), (ii) and (iii), the quadri-tangent ellipses are easily identified using the symmetries of these polynomials and the system of equations (34).

The ellipses quadri-tangent to $x(x^2 + 3y^2)$ are defined by matrices of the form $H_\lambda = \mathrm{diag}(\lambda, \frac{4+\lambda^3}{3\lambda^2})$, where $0 < \lambda \leq 2$. Note that $\det H_\lambda$ is decreasing on $(0, 2^{\frac{1}{3}}]$ and increasing on $[2^{\frac{1}{3}}, 2]$. Given π with $\mathrm{disc}\,\pi < 0$, the optimization problem (36), therefore becomes

$$\min_\lambda \{\det H_\lambda \ : \ (\phi_\pi^{-1})^{\mathrm{T}} H_\lambda \phi_\pi^{-1} \geq \alpha\,\mathrm{Id}\}.$$

If the constraint is met for $\lambda = 2^{1/3}$, we obtain $E_{\pi,\alpha} = E_\pi$ and therefore $\alpha \leq \beta_\pi$. Otherwise, using the monotonicity of $\lambda \mapsto \det H_\lambda$ on each side of its minimum $2^{\frac{1}{3}}$ we see that the matrix $H_\lambda - \alpha\phi_\pi^{\mathrm{T}}\phi_\pi$ must be singular. Taking the determinant, we obtain an equation of degree 4 from which λ can be computed, and this concludes the proof of part 3 (i).

The ellipses quadri-tangent to $x(x^2-3y^2)$ are defined by $H_{\lambda,V} = V^{\mathrm{T}}\,\mathrm{diag}(\lambda, \frac{4-\lambda^3}{3\lambda^2})V$, where $0 < \lambda \leq 1$ and V is a rotation by $0, 60$ or 120 degrees. Since $\det H_{\lambda,V}$ is a decreasing function of λ on $(0, 1]$, we can apply the same reasoning as above to polynomials π such that $\mathrm{disc}\,\pi > 0$. This concludes the proof of part 3 (ii).

Last, the ellipses quadri-tangent to xy^2 are defined by $H_\lambda = \mathrm{diag}(\lambda, \frac{4}{27\lambda^2})$, $\lambda > 0$. The determinant is a decreasing function of λ, with lower bound 0 as $\lambda \to \infty$, and the same reasoning applies again hence concluding the proof of part 3 (iii).

B Proof of non existence property in Theorem 6.1

Let (m, d) be an incompatible pair. We know from Lemma 6.1 that there exists $\pi_0 \in \mathbb{H}_{m,d}$ such that $K_{m,d}(\pi_0) > 0$ and $Q(\pi_0) = 0$ for all invariant polynomial $Q \in \mathbb{I}_{m,d}$.

We assume for contradiction that a polynomial $\mathbf{K}$ satisfies inequalities (44). Up to replacing $\mathbf{K}$ with $\mathbf{K}^{2d}$, we can assume that $\mathbf{K}$ takes non negative values on $\mathbb{H}_{m,d}$ and that $\mu = \frac{mr}{d}$ is an integer. The rest of this proof consists in showing that $\mathbf{K}$ *needs to be* an invariant polynomial, thus leading to a contradiction since we would then have $\mathbf{K}(\pi_0) = 0$. For this purpose we derive from inequalities (44), and from the invariance of $K_{m,d}$ with respect to changes of variables, the inequalities

$$C^{-2r}(\det \phi)^{\mu}\mathbf{K}(\pi) \leq \mathbf{K}(\pi \circ \phi) \leq C^{2r}(\det \phi)^{\mu}\mathbf{K}(\pi), \tag{45}$$

where C is the constant appearing in inequalities (44). We regard the function $\mathbf{Q}(\pi, \phi) = \mathbf{K}(\pi \circ \phi)$ as a polynomial on the vector space $V = \mathbb{H}_{m,d} \times M_d$, where M_d denotes the space of $d \times d$ matrices, and observe that it vanishes on the *hypersurface* $V_{\det} := \{(\pi, \phi) \in V \ : \ \det \phi = 0\}$. Since $\phi \mapsto \det(\phi)$ is an irreducible polynomial, as shown in [5], it follows that $\mathbf{Q}(\pi, \phi) = (\det \phi)\mathbf{Q}_1(\pi, \phi)$ for some polynomial $\mathbf{Q}_1$ on V. Injecting this expression in inequality (45) we obtain that $\mathbf{Q}_1(\pi, \phi)$ also vanishes on the hypersurface $V_{\det}$ and the argument can be repeated. By induction we eventually obtain a polynomial $\widehat{\mathbf{K}}$ on V such that $\mathbf{K}(\pi \circ \phi) = (\det \phi)^{\mu}\widehat{\mathbf{K}}(\pi, \phi)$. It follows from inequality (45) that for all $(\pi, \phi) \in V$

$$C^{-2r}\mathbf{K}(\pi) \leq \widehat{\mathbf{K}}(\pi, \phi) \leq C^{2r}\mathbf{K}(\pi).$$

This implies that $\widehat{\mathbf{K}}(\pi, \phi)$ does not depend on ϕ. Otherwise, since it is a polynomial, we could find $\pi_1 \in H_{m,d}$ and a sequence $\phi_n \in M_d$ such that $|\widehat{\mathbf{K}}(\pi_1, \phi_n)| \to \infty$. Therefore

$$\mathbf{K}(\pi \circ \phi) = (\det \phi)^{\mu}\widehat{\mathbf{K}}(\pi, \phi) = (\det \phi)^{\mu}\widehat{\mathbf{K}}(\pi, \mathrm{Id}) = (\det \phi)^{\mu}\mathbf{K}(\pi).$$

This establishes the invariance property of $\mathbf{K}$, in contradiction with our first argument, and concludes the proof.

Acknowledgement

I am grateful to to professor Nira Dyn for her invitation in Tel Aviv University where this work was conceived, and to my PhD advisor Albert Cohen for his support in its elaboration.

References

[1] F. Alauzet and P.J. Frey, *Anisotropic mesh adaptation for CFD computations*, Comput. Methods Appl. Mech. Engrg. 194, 5068-5082, 2005.

[2] T. Apel, *Anisotropic finite elements: Local estimates and applications*, Advances in Numerical Mathematics, Teubner, Stuttgart, 1999.

[3] V. Babenko, Y. Babenko, A. Ligun and A. Shumeiko, *On Asymptotical Behavior of the Optimal Linear Spline Interpolation Error of C^2 Functions*, East J. Approx. 12(1), 71–101, 2006.

[4] Yuliya Babenko, *Asymptotically Optimal Triangulations and Exact Asymptotics for the Optimal L^2-Error for Linear Spline Interpolation of C^2 Functions*, submitted.

[5] M. Bocher, *Introduction to Higher Algebra*, Courier Dover Publications, 2004 ISBN 0486495701, 9780486495705

[6] J-D. Boissonnat, C. Wormser and M. Yvinec. *Locally uniform anisotropic meshing.* To appear at the next Symposium on Computational Geometry, june 2008 (SOCG 2008)

[7] Sebastien Bougleux and Gabriel Peyré and Laurent D. Cohen. *Anisotropic Geodesics for Perceptual Grouping and Domain Meshing.* Proc. tenth European Conference on Computer Vision (ECCV'08), Marseille, France, October 12-18, 2008.

[8] Y. Bourgault, M. Picasso, F. Alauzet and A. Loseille, *On the use of anisotropic error estimators for the adaptative solution of 3-D inviscid compressible flows,* Int. J. Numer. Meth. Fluids. [Preprint]

[9] W. Cao. *An interpolation error estimate on anisotropic meshes in $\mathbb{R}^n$ and optimal metrics for mesh refinement.* SIAM J. Numer. Anal. 45 no. 6, 2368–2391, 2007.

[10] W. Cao, *Anisotropic measure of third order derivatives and the quadratic interpolation error on triangular elements*, SIAM J.Sci.Comp. 29(2007), 756-781.

[11] W. Cao. *An interpolation error estimate in $\mathbb{R}^2$ based on the anisotropic measures of higher order derivatives.* Math. Comp. 77, 265-286, 2008.

[12] L. Chen, P. Sun and J. Xu, *Optimal anisotropic meshes for minimizing interpolation error in L^p-norm*, Math. of Comp. 76, 179–204, 2007.

[13] A. Cohen, N. Dyn, F. Hecht and J.-M. Mirebeau, *Adaptive multiresolution analysis based on anisotropic triangulations*, preprint, Laboratoire J.-L.Lions, submitted 2008.

[14] A. Cohen, J.-M. Mirebeau, *Greedy bisection generates optimally adapted triangulations*, preprint, Laboratoire J.-L.Lions, submitted 2008.

[15] R. DeVore, *Nonlinear approximation*, Acta Numerica 51-150, 1998

[16] J. Dixmier, *Quelques aspects de la théorie des invariants*, Gazette des Mathématiciens, vol. 43, pp. 39-64, January 1990.

[17] P.J. Frey and P.L. George, *Mesh generation. Application to finite elements*, Second edition. ISTE, London; John Wiley & Sons, Inc., Hoboken, NJ, 2008.

[18] R. Hartshorne, *Algebraic Geometry.* New York: Springer-Verlag, 1999.

[19] D. Hilbert, *Theory of algebraic invariants*, Translated by R. C. Laubenbacher, Cambridge University Press, 1993.

[20] F. Labelle and J. R. Shewchuk, *Anisotropic Voronoi Diagrams and Guaranteed-Quality Anisotropic Mesh Generation*, Proceedings of the Nineteenth AnnualSymposium on Computational Geometry, 191-200, 2003.

[21] S. Lang, *Algebra*, Lang, Serge (2004), Algebra, Graduate Texts in Mathematics, 211 (Corrected fourth printing, revised third ed.), New York: Springer-Verlag, ISBN 978-0-387-95385-4

[22] P.J. Olver, G. Sapiro and A. Tannenbaum, *Affine invariant detection; edge maps, anisotropic diffusion and active contours*, Acta Applicandae Mathematicae 59, 45-77, 1999.

[23] G. Salmon, *Higher plane curves*, third edition, 1879: http://www.archive.org/details/117724690

[24] J. R. ShewChuk, *What is a good linear finite element*: www.cs.berkeley.edu/˜jrs/papers/elemj.pdf

[25] A mathematica code for the map $\pi \mapsto h_\pi$: www.ann.jussieu.fr/˜mirebeau/

[26] The 2-d anisotropic mesh generator BAMG: http://www.freefem.org/ff++/ (included in the FreeFem++ software)

[27] A 3-d anisotropic mesh generator: http://www.math.u-bordeaux1.fr/˜dobj/logiciels/ mmg3d.php

Jean-Marie Mirebeau
UPMC Univ Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France
CNRS, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France
mirebeau@ann.jussieu.fr